

Europe's parliament in the doldrums

This week's re-election of the European Parliament will have been a futile exercise unless the newly elected can find a role for themselves; the best bet will be to engineer the intergovernmental conference planned for 1996.

THIS week's elections of members of the European Parliament (MEPs) are at once a reminder that the European Union (EU) is a reality and that the reality falls a long way short of the ideal. Between today (9 June) and next Sunday (12 June), all 12 member states are required to re-elect members to the Strasbourg parliament (which is sometimes to be found in Brussels and sometimes, for some of its functions, in Luxembourg). The representation afforded by these arrangements is far from equal. Small countries have more than a fair share of MEPs, for example. The representation is also variable in character. European law allows national governments to organize European elections as they see fit. Some have settled for proportional representation, others for constituency elections in which the winning party's candidate is elected. There is a general complaint throughout Europe that MEPs, divorced as they are from national politics, cannot represent their notional constituents in any meaningful way. MEPs themselves complain that they have too little influence on EU affairs.

Yet the European Parliament is supposed to be a crucial institution of the EU. How can its reputation, among those who belong to it as well as those who vote for it, be so low? The simple truth is that it has hitherto been kept deliberately in that condition by most member governments, who have their say on EU affairs through the European Council, and would not relish a system in which their decisions were qualified by a parliament in some far-off European city. Of course, the Maastricht Treaty, now in force, gives the parliament more power to scrutinize and to comment on draft European legislation; will that give frustrated MEPs a greater sense of usefulness? It seems improbable — the parliament has too little coherence around common principles and goals to be both an effective and a constructive critic of legislation. That could change, but only with time.

So why is the whole of Europe spending all this time and energy electing MEPs? The explanation is to be found in the words "ever closer union" found in the preamble to the Maastricht Treaty. Europe's parliament is a parliament waiting for a state of affairs that has not yet arrived. The question of how it should occupy itself in the interval is then more readily answered. In its own long-term interests, it should make the cause of European union more coherent and persuasive. The battle over Maastricht and its ratification was so long and bruising for so many member governments precisely because the European Communities (as the EU then was) had no effective way of explaining centrally what

Maastricht was about. (National governments were eager to put their own gloss on it, the European Commission was inevitably constrained by the fear that pronouncements from Brussels would be denounced as internal political interference by member states.)

But how can the parliament advance the cause of "ever closer union" when many of those standing for election this week are opposed to the idea? Admirably, at least in one important respect. In retrospect, one of the serious errors at Maastricht was that governments put their names to a treaty sewn together by their own officials with such contrived intricacy that politicians could not easily pick it apart. The result is that many of its proposals were detached from political and even institutional reality. (The project for European Monetary Union, one of the centrepieces of Maastricht, had come unstuck within nine months of ratification.) It would have been particularly valuable if sceptical members of the European Parliament had some say long before the treaty came up for signature.

That is the cue to which the newly elected parliament should jump. Member states (which may then be 16-strong) are committed to another intergovernmental conference in 1996, when the agenda will explore new forms of political cooperation. Already several contentious issues have identified themselves, from immigration policy to the question of what might have been done about Bosnia and Herzegovina. On past form, officials will already be beaver away at the design of compromises for 1996. The new parliament should seek to force its way into that process. Sceptical opinions of what is proposed would be especially valuable, if only as warnings of trouble ahead. But how would that advance the European cause? It is more important that the product of the next intergovernmental conference should be a practical and easily acceptable treaty than that it should be an ambitious one. The European Parliament may lack some aspects of legitimacy, but it has enough to ensure that. □

Is North Korea nuclear?

The rest of the world cannot let North Korea go nuclear, but needs guile to prevent it.

WITH Iraq now quiet, North Korea is the most troublesome signatory of the Nuclear Non-Proliferation Treaty (NPT). It seems generally agreed that North Korea declined to allow



safeguards inspectors from the International Atomic Energy Agency (IAEA), on their visit at the end of last month, to inspect fuel rods discharged from a working reactor. The point is important because even external inspection can tell whether fuel rods have been handled so as to maximize the amount of fissile plutonium (^{239}Pu) or whether, instead, they have been used to extract as much power as possible (in which case ^{240}Pu predominates). And there are two explanations of North Korea's refusal; either North Korea has been making military plutonium and wishes not to be found out, or it is innocent, but wishes by its refusal to create the impression that it has a nuclear bomb or even several.

Whatever the truth, North Korea's refusal is an infringement of its obligations under the NPT. Moreover, the timing of this latest trouble for the NPT could hardly be worse. Just about a year from now, the United Nations will be convening the review conference of the treaty members at which it is hoped to win general agreement for the continuation of the treaty. It will hardly be a happy augury for the outcome of that meeting if one of the treaty's signatories is, at the same time, being subjected to the United Nations sanctions that the United States seems ready to demand from the Security Council. There are two further complications. China, whose position on North Korea as a stripling nuclear power is far from clear, may yet veto joint action by the United Nations. And North Korea's threat to regard sanctions as "an act of war" cannot altogether be dismissed as empty.

The most serious worry, which will not quickly go away, is that North Korea is now more isolated from the rest of the world than it has ever been. It is a backwater in a region of almost spectacular economic growth. Even China, its ally in the Korean War of the 1950s and its doctrinal supporter until recently, has opted for economic improvement. South Korea, the other side of the demilitarized zone surrounding the armistice line (there is no peace treaty), must seem from Pyongyang to be a constant reminder of how different things might have been in the North. Yet, incredible as it may seem, there is something in the view that the government of North Korea is crucially, if partially, sustained by remittances of cash from the Korean community in rich Japan.

Given all the difficulties of the region, the safe and even civilized solution is to make it palatable for North Korea to end its isolation from the rest of the world. One obvious difficulty is that it is a Marxist dictatorship of the old-fashioned kind. Another is that its government is hardly on speaking terms with most others. The exception is China, which is why there are the strongest possible reasons for asking that China should take the lead in making North Korea honour its treaty obligations and then persuading it to join the world around it. To risk a Chinese veto of a sanctions proposal in the Security Council would have the opposite effect.

Why should North Korea be offered kid-glove treatment it clearly does not deserve? The truth is that the maintenance and, ideally, strengthening of the NPT is a more important goal than the equitable distribution of big-power punishments for bad behaviour. It is not as if North Korea is the only problem to be dealt with before next year's conference.

Apart from the familiar issues of the Middle East and the Indian subcontinent, the eventual fate of fissile material from Russia and the other ex-Soviet republics, the international community has to devise a regime for keeping out of harm's way huge quantities of other people's fissile material. It has just a year in which to do the job. The danger is that the time available will be frittered away in bothering about North Korea and its malevolent intentions. That would be to throw out the baby with the bathwater. □

Charitable research

British charities will not, after all, have to vouch for the quality of what their dependent researchers publish.

THE British Charity Commission has wisely backed away from a proposal it aired two months ago that seemed to require charities making research grants to vouch for the authenticity of publications arising from work carried out with their funds. The occasion was the continuing fuss about an article in *The Lancet* in 1990 pouring cold water on claims by the Bristol Cancer Health Centre that alternatives to conventional treatments for some types of cancer had proved beneficial (see *Nature* 367, 100; 1994). The published account of the research is now acknowledged to have been misleading, to say the least. The commissioners' first reaction was to say that the charities supporting the research should have taken responsibility for accounts of it eventually published.

That, of course, would be unworkable, as seems now to be recognized. In a set of guidelines put out last week, the Charity Commission recognizes that charities should stand in relation to research just as do public grant-making agencies in Britain. In other words, there are contractual relationships between the source of funds and both the recipient researcher and the researcher's institution. But it is for the institution, which is usually the researcher's employer, to be responsible for the proper conduct of the research, and for journals to ensure the quality of what is published. The Charity Commission promises a consultation paper on these and other issues.

So far, so good. Indeed, there are probably many British charities that will benefit from being told in plain language that research proposals should be sent to referees when the trustees themselves are not expert in the field concerned. But it must be hoped that the version of the guidelines eventually agreed will be appropriately flexible. The great value of independent charities in support of research is that they are often able and willing to back projects that do not survive the solemn processes of peer review, or which are structurally unusual in being devised to persuade others with longer purses to spend even more. Charities are also willing (rightly) to back what is sometimes called 'action research', perhaps mounting projects that appear to be socially valuable well in advance of the point at which they can be assured of success. What the Charity Commission now has in mind as guidance is generally correct, but it will be unfortunate if its eventual guidelines are too prescriptive. □

Concern over 'invisible problem' of HIV blood in developing countries

Paris. Almost ten years after the industrialized world began to screen all blood used in transfusions for human immunodeficiency virus (HIV), as many as one in ten seropositives in developing countries are still being infected through this route.

Aware of the continuing problem, the World Health Organisation (WHO) last week approved the setting up of a new blood safety unit. But some observers fear that political wrangling within WHO, combined with cutbacks at the United Nations (UN), will deny the unit the funds and staff that it needs to focus both international attention and resources on the problem of blood safety.

The screening of blood for HIV is patchy in the developing world. Even in Côte d'Ivoire, for example, which has one of the most advanced systems, such screening is routine only in the capital Abidjan, and in two other cities.

A combination of the lack of screening with high levels of infected donors turns transfusion into a form of Russian roulette. A survey carried out by WHO in 1990, for example, showed that centres in Kinshasa, Zaïre, screened fewer than a quarter of blood units for HIV, even though about five per cent of donors were seropositive.

Receiving a transfusion of HIV-contaminated blood carries a roughly 95 per cent risk of HIV infection, compared with a 0.1 to 1.0 per cent risk through sexual transmission. "Endorsing risky transfusion is as morally acceptable as saying that it's okay for an antibiotic to kill one-third of patients", says one French scientist.

But as a result, transfusion still accounts for 5 to 10 per cent of HIV infections worldwide, according to Jean Emmanuel, of the WHO Global Programme on AIDS. In some developing countries, the figure is even higher, he says.

Despite this, blood safety has yet to be given priority by the international community. "There is a sort of agreed invisibility to the problem," says Daniel Tarantalo, director of the International AIDS Programme at the Harvard School of Public Health. "Countries are embarrassed to admit their blood is unsafe, while the international agencies don't dare expose individual countries, as they would be held responsible for not providing enough of the right support."

Some WHO officials also feel that money is better spent on preventing sexual transmission of HIV. Funding for HIV screening of blood is being phased out of WHO's Global Programme on AIDS, for example. "There is generally less and less money

available for screening," says Souleymane M'Boup, head of the laboratory of bacteriology and virology at the Aristide Le Dantec University Hospital in Dakar, Senegal.

Similarly, a Global Blood Safety Initiative, announced by WHO five years ago, never took off. According to one WHO official, its credibility was destroyed by underfunding and understaffing. "GBSI was a valid approach," says Tarantalo. "But

unit is a victory for Fernando Antezana, who has fought for greater emphasis on blood safety since he became assistant director-general of WHO earlier this year.

Many applaud his stand; "we are now on the brink of getting our act together," says one WHO official. But others remain concerned that Antezana may not be able to win sufficient funding for the unit, particularly at a time when programmes and posts are being cut back throughout the UN.

If Antezana does obtain the money he needs, the practical tasks he faces will still nonetheless be formidable. A lack of HIV screening is usually only a symptom of an already impoverished transfusion service, and these services need to be revamped before screening systems can work well.

Senegal, for example, has one of the few comprehensive HIV screening programmes in Africa, primarily because it had an efficient blood service in operation before the dangers of the virus were known, says M'Boup.

In contrast, the WHO survey in Kinshasa in 1990 showed that half the blood centres had no refrigeration, that 90 per cent lacked facilities for testing donor/recipient compatibility and that bacterial contamination of blood was "common".

One major plank of any aid programme will be the building of production facilities for blood substitutes, says Jean-Jacques Fournel, director of blood transfusion at the Pitié-Salpêtrière hospital in Paris. Many developing countries use blood unnecessarily, he says, as it is cheaper than more appropriate alternatives, such as saline solutions or colloids, which have to be imported.

In addition, Fournel says that well-equipped screening programmes often founder because they fail to take local culture sufficiently into account. In particular, procedures are often not followed because of a lack of proper supervision and training. Africans often prefer to make decisions by consensus rather than giving orders to one another, he says. "It may sound reactionary, but you need foreign supervision to get things done."

Declan Butler

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Blood donation in Africa: funds for screening are still lacking.

inter-institutional tensions were insurmountable."

One senior WHO official admits that what he describes as a "lack of coordination" among international organizations is a major problem. For example, UN agencies have, despite almost a year of negotiation, yet to agree how to supply HIV test kits to Burma, where two-thirds of blood goes unscreened. "Institutional mandates are unclear, and each agency has competing priorities within its budgets," says Tarantalo.

Another problem has been straightforward bungling by aid agencies. For example, screening equipment was sent to Ethiopia without local laboratory staff having been trained to use it.

Similarly, WHO has sent reagents with a two-month shelf life to countries where they are not being used for six months. Reagents requiring refrigeration are often left in the sun waiting for customs clearance.

WHO's approval of the new blood safety

New setback for India's nuclear programme

New Delhi. India's nuclear power programme, already truncated due to financial cutbacks, suffered further last month when a 1500-tonne concrete dome of a reactor containment building under construction at Kaiga in Karnataka state partly collapsed.

No radioactivity was released, as nuclear fuel had not been transferred to the building,

and the 14 workers injured required no more than first-aid treatment. However the incident, coming less than a year after a damaging fire at another atomic plant in Narora in Uttar Pradesh, has given fresh ammunition to the state's environmental groups, which are opposed to the plant's location in virgin rain forest. □

Doubts greet violence-research funding plea

Washington. A substantial increase in the sums allocated in the United States to violence-related research has been recommended by an expert panel in a report last week to Harold Varmus, the director of the National Institutes of Health (NIH).

The panel points out that violence should be considered a major public health problem, as it is the second leading cause of death among American youth. And African American men are seven times more likely than other Americans to be the victims of homicide, it says.

But both Varmus and the members of his advisory committee, at a meeting at which the panel's report was made public, expressed scepticism about the value of such an increase without a clear plan for how the extra money would be spent.

The expert panel was established last year by Bernadine Healy, Varmus's predecessor as NIH director, in response to controversy surrounding a grant for a conference on genetic factors and crime, made by the National Center for Human Genome Research to the University of Maryland.

A brochure produced by the university about the conference angered members of the African American community, particularly at Howard University in Washington, DC. The main complaint was that inadequate attention was being paid to the many factors apart from genetics — for example

poverty — that are causes of violence.

Healy suspended the grant for the conference on the grounds that the controversy would not allow a proper scientific debate to be held (a decision that was later overturned

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Despite decades of research into its causes, violence remains on the increase in the US.

by an appeals board of the Public Health Service). At the same time, she decided that the NIH should take a close look at all violence-related research carried out in its institutes.

Although the controversial origins of the report were barely touched on at last week's meeting, Don Wood, professor of neurology at Howard University Hospital, commented on the absence from its 19 recommendations of any explicit statement identifying a particular area of questioning as a "no go" area. "I think that this should be explicitly

stated," he said.

Thomas Murray, co-chair of the panel and director of the Center for Biomedical Ethics at Case Western Reserve University, in Cleveland, Ohio, preferred to emphasize his proposal that funding should double — or even triple — from its current level of \$58 million a year. He also called for interdisciplinary research and the introduction of a new peer-review panel qualified to evaluate such research.

But several members of the advisory committee appeared unimpressed by the report's conclusions. Among them was Joshua Lederberg, of the Rockefeller University in New York, who claimed "an intense sense of *déjà vu*". According to Lederberg he could have heard the same conclusions 25 years ago when he sat as a member of an advisory panel to the National Institute of Mental Health. "We need to know what we have learned to date," he said. "We need some focus, vision and working hypotheses in this area."

Other members of the advisory committee felt it should have paid more attention to basic biomedical research. Varmus himself seemed less than enthusiastic about its findings. He questioned why the report contained no assessment of the achievements of violence-related research to date, or details of how the extra money should be spent.

Helen Gavaghan

... as NIH tightens up on academic-industry deals

Washington. Close scrutiny should be given to industrial research contracts signed by any academic research group if the contract represents more than 20 per cent of that group's overall budget, according to proposed conflict-of-interest guidelines published last week.

Other situations that should trigger such scrutiny include those where financial support from industry exceeds \$5 million in one year — or \$50 million in total — and where the licensing rights cover all, or a substantial proportion, of the technologies developed from the group's research in a particular field.

Furthermore, an agreement lasting for five or more years should also receive close attention, according to the guidelines on academic-industry agreements presented

last week to Harold Varmus, the director of the National Institutes of Health, and now being circulated for comment in the academic community.

The proposed guidelines have been drawn up by an *ad hoc* advisory panel to the NIH in the wake of last year's public controversy over an agreement signed between the Scripps Research Institute in San Diego, California, and the Swiss company Sandoz Pharmaceutical.

In principle, such deals have long been encouraged by the government, particularly after the passage of the Bayh-Dole Act in 1980, which sought to boost the transfer of research results from universities to the marketplace. This act allows universities to retain title to inventions resulting from federally funded research, and to licence the inventions to industry for development.

There are now some 7,000 licensing agreements of various kinds with industry. Most have proceeded without controversy. But the Scripps-Sandoz deal raised congressional hackles because it effectively gave Sandoz exclusive access to all publicly funded research carried out at Scripps.

Following lengthy negotiations with the

legal department of the NIH, the two parties have agreed to modify their agreements. Originally Sandoz was to have paid \$300 million over 10 years for access to about \$1-billion worth of federally funded research at Scripps. The company will now pay \$20 million a year for five years, and will in return have exclusive rights to only 47 per cent of the institution's research.

Responding to criticism that the original deal was too tightly tied to the interests of a single large company, Scripps is also establishing a programme to encourage technology transfer to small businesses, which the NIH will monitor annually.

Last month, the NIH sent to Scripps what it hopes will be the last letter in the controversy. In it, Varmus wrote that, given the institution's cooperation in making research available to small businesses, the NIH hoped to reach an agreement that would obviate the need "to limit Scripps' right to federally funded inventions".

According to one NIH official, the letter is intended as a shot across the bows, indicating to other large academic institutions that there are limits to what the NIH is prepared to tolerate.

H. G.

Italian science

Italy's new minister for universities and research is Stefano Podestà, and not Bodestà, as published in our issue of 19 May (369, 175; 1994). Also, the proposals put forward by his predecessor, Umberto Colombo, would increase spending on research and development from 1.4 per cent to 1.6 per cent of the gross national product, not 1.6 per cent to 2.0 per cent as printed.

War of words over future of Berlin teaching hospitals

Munich. The Free University of Berlin seems likely to lose its fight to stop one of its two teaching hospitals being transferred to Humboldt University in the east of the city. Last week, the Christian Democrat faction of Berlin's city government proposed a bill that would lead to a transfer next April.

But the Free University, in the former West Berlin, is fighting to retain at least some control over the hospital. The dispute has become a bitter power struggle between the two universities, fanned by concern over the financial cuts being imposed on institutions in the west of the city to support the strengthening of those in the east.

Reunification of the city after the fall of the Berlin wall in 1989 quickly turned the west half from a rich and privileged (if isolated) city, heavily subsidized by the federal government in Bonn, into one marked by social unrest, economic collapse and political turmoil.

The highest priority of the Berlin government is now to save money, and the ministry of science and research has been given the urgent task of rationalizing the city's three universities. Most of the required department mergers have now been achieved (see *Nature* 363, 198; 1993). But the issue of the teaching hospitals has continued to fester.

In March, the ministry decided to recommend the merger of two such hospitals — the Charité, belonging to Humboldt University, and the Rudolf Virchow Hospital, belonging to the Free University — under the management of Humboldt University. (The Free University's Steglitz Hospital, with about 1,200 beds, would remain where it is.)

The ministry wants the Charité to eliminate 150 beds and the Rudolf Virchow 200, leaving them with 1,200 and 1,150 beds respectively, and the total number of staff in both hospitals to be reduced by natural wastage over the next few years (a target number has not yet been defined).

It also wants an estimated DM1.6 billion (US\$960,000) for building repairs and modernization of the Charité to be reduced by sharing facilities and equipment with the Rudolf Virchow hospital, whose new buildings are only 2 km away.

Complying with Humboldt University's wish to complete the post-reunification reassessment of all staff before merging departments, the ministry suggests as a first step the transfer of the management of the Virchow hospital to Humboldt University in April next year.

The Christian Democrat faction of Berlin's coalition government last week put forward a draft bill based on the ministry's suggestions. Details are being discussed with

its coalition partners, and the bill is expected to go to parliament for approval in the next few weeks.

But the proposals have not gone down well with the Free University, unhappy about losing control over a teaching hospital in which it has invested nearly DM2 billion in the past 10 years.

Vice-president Peter Gaegtens, a physiologist, says that the decision to leave the Free University with only a single hospital would put the university at an unfair scientific disadvantage compared to Humboldt, as the loss of the Virchow hospital would result in preclinical teaching and research resources at the university being cut by two-thirds.

Gaegtens says it is unclear how the ministry's proposals would save money. Together with the president, Johann Gerlach, he wants the ministry to produce a detailed financial plan showing that, by giving up its hospital, the university will be saving Berlin a considerable amount of money.

Furthermore, if the ministry's only motivation is to save money, it should be proposing an immediate merger, rather than waiting for the Charité to reappoint its professors, as this will reduce its room for manoeuvre during the later merger, says Gaegtens.

In reply, the ministry argues that savings will come from the proposed cuts in beds and staff, but says it cannot provide exact figures until negotiations are complete.

The Free University has put forward counter-suggestions, such as a merger of all three teaching hospitals into a single medical school independent of the three universities, or even merging the Humboldt and Free universities along the lines of the multi-sited University of London. But the ministry does not consider either suggestion to be practical in the present situation in Berlin.

Temper is running high. Gerlach and Gaegtens have attacked the ministry's proposals for merging the hospitals as "dishonest, deceitful and irresponsible". Manfred Erhardt, Berlin's science minister, has hit back by saying that they are using colourful language to hide the weaknesses of their arguments, dismissing the critics as "two commanders condemning an enemy to whom most of their troops have already defected".

The Free University is unlikely to succeed in its fight, as this would disrupt overall plans for restructuring the university system in Berlin, and bringing Humboldt up to the academic standards in the west. But the bitterness of the conflict, given the general maelstrom of changes in Germany's future capital, is trying the nerves of all involved, while politeness and tolerance have been relegated to the past.

Alison Abbott

US unveils details of \$750 million technology plan

Washington. The National Institute of Standards and Technology (NIST), part of the US Department of Commerce's Technology Administration, last week fleshed out its proposal to provide focused support on a cost-sharing basis to industrial research and development projects in five key areas of technology.

Areas chosen for this five-year, \$745-million Advanced Technology Program (ATP) initiative include DNA-based diagnostics, the manufacture of composite structures, the development of an information infrastructure for the health-care sector, computer-integrated manufacturing for electronics, and the development of component-based software.

For the current fiscal year, the total budget for the ATP is just under \$200 million, an almost three-fold increase over last year. The ATP is due to receive further major increases under President Bill Clinton's Economic Plan which, if approved by Congress, would increase its budget to almost \$750 million by 1997.

The goal of the ATP is to support new technologies with a high level of technical risk, but with significant commercial potential, that private and public investors might be reluctant to back. Proposals are invited from single companies or joint ventures consisting of at least two for-profit companies.

Although the ATP's focus on private companies is what distinguishes it from other forms of federal funding, universities, government laboratories and other non-profit organizations can still apply for support, either as subcontractors to single companies, or to a joint venture. Such organizations can receive payment from companies in recognition of their contribution. But title to any patents resulting from ATP-sponsored research must be held by a company incorporated in the United States.

Some are finding this legislatively-mandated provision particularly hard to swallow. Lita Nelsen, director of the Technology Licensing Office at the Massachusetts Institute of Technology (MIT), says that it does not believe the restriction is appropriate, "nor do we believe it was the intent of Congress". A technical amendment clarifying the original intention of the law is now pending in Congress. But until such time as the law is changed, Nelsen says MIT remains reluctant to enter into such agreements.

The \$145-million "Tools for DNA Diagnostics" competition invites proposals that will lead to development of new diagnostic methods, instrumentation and data-management tools designed to speed up DNA analyses and drive down costs by an order of magnitude or more.

Diane Gershon

Head of Soros fund rejects complaints over new grants

Moscow. The International Science Foundation (ISF) has allocated its second round of awards to scientists in the former Soviet Union. The grants, worth a total of US\$45 million, follow the first round of small emergency awards from the ISF, designed to fund successful applicants until mid-1995.

But whereas the distribution of the initial short-term grants was greeted with general approval, that of the second batch has left the ISF, set up with funding from the Hungarian-born financier George Soros, swamped with complaints. Some scientists, for example, are unhappy at the large number of applications that were turned down (only 3,000 grants were approved out of a total of 15,000).

Others are complaining that they received significantly less than they asked for, and claim that applicants were rejected even though they scored higher marks in the evaluation than successful colleagues.

But Alexander Goldfarb, director of the ISF's Washington office, rejects any charge of unfairness. He argues that the complex reviewing procedures used by the ISF — a mixture of peer review and assessment by expert panels — may well explain the high level of discontent.

Peer review resulted in more objective evaluations, he said, but the reviewer was not able to compare competing applications and rank them according to priority for funding. "The members of the expert panel are able to make such a comparison, even though their final joint decision may be by far more subjective," says Goldfarb.

According to Goldfarb, Western reviewers were often reluctant to give bad marks to impoverished colleagues in Russia and elsewhere. "If we were to rely just on those assessments, the average grant would have to be reduced to between \$4,000 and \$5,000, which is not enough even for food in Russia." The more stringent selection procedure adopted by the ISF allowed it to increase the average grant to \$17,000.

Despite the many complaints, the geographical distribution of the grants showed a similar pattern both to that of the earlier emergency grants and to that of papers published in scientific journals. As a result, Alexander Bratus, scientific secretary of one of ISF's panels, claims that the results of the long-term grant competition accurately reflect the situation in Russia.

Some scientists believe that panel members came under strong pressure from the applicants and other interested parties. Goldfarb admits that, despite precautions such as discouragement of lobbying and immediate intervention where it was suspected, some of this pressure may have filtered through.

Vladimir Pokrovsky

Energy and biotechnology top Euro-Parliament's agenda

Paris. When the European Union's (EU)'s 269 million voters go to the polls this week to elect a new European Parliament, they will be voting mainly on national issues, and most will be indifferent to science.

But the representatives they elect will have an increasing say in shaping the future of EU research. Although the parliament cannot make laws, successive treaties have given it increasing influence.

In particular, the Maastricht Treaty, which came into force last year, gives the parliament a veto on the budget of research under the Framework programme, which it insists carries with it the right to help shape EU research policy.

The parliament is supposed to compensate for the 'democratic deficit' of the two other main EU institutions, the European Commission (EC) and the Council of Ministers. The Maastricht agreement, however, has not been in force long enough for parliament to know how effective its veto and other new powers will be.

The first major test of the parliament's new powers will be a debate on an EU energy policy, which the commission is scheduled to propose during the six-month German presidency beginning on 1 July. Liberalization of Europe's energy monopolies will dominate the debate. But the parliament is also likely to seek a decentralization of energy production, and greater emphasis on renewable energy sources.

Past policies also suggest that, when it comes to discussing the fifth Framework programme, parliament will seek more funding for life sciences, environment, renewable energy and new industrial technologies, and less for costly information and telecommunications programmes.

Funding for fusion research — which receives about ECU1 billion within the fourth Framework — may also experience more difficulty next time round. Some members refer to fusion as "the emperor with no clothes", and Detlev Sanland, who once tabled a motion against fusion research, is said to be a possible candidate to head the parliament's powerful budget committee.

In the immediate future, biotechnology will probably be the major research topic on the parliament's agenda. One of its first tasks, for example, will be to approve the content of the biotechnology and biomed-

cal programmes within the fourth Framework programme.

Another more controversial task will be to reach final approval of a new directive standardizing national laws on the patenting of life. After a voting muddle last month, parliament only adopted three amendments

before attendance fell below the 260 required.

Under the new rules of the Maastricht agreement, approval now requires parliament and the Council of Ministers to agree on a single text in a 'conciliation committee'. Parliament, feeling short-changed, is likely to hold out for inclusion of some of the amendments it was unable to vote through last month.

Indeed, biotechnology has risen up the EU agenda following the release earlier this year of a white paper on competitiveness,

growth and unemployment, by Jacques Delors, the president of the commission. This urged the EU to develop a strategy in biotechnology as a matter of urgency.

SAGB, a Brussels-based industry lobby group, has leapt on this, recently criticizing the EC for not doing enough to create "structural, regulatory, and fiscal incentives" to encourage investment in biotechnology.

SAGB also says that proposals made by the commission last week to reduce regulatory requirements for using genetically modified organisms (GMOs) — especially those considered low-risk — do not go far enough. But the parliamentary Green group is furious with the proposals, which they claim would lead to an unacceptable loosening of safety controls.

Claude Desama, president of the parliament's energy, research and technology committee (ERTC), says he "regrets" that the commission ignored his request for discussions before the proposals were issued. But he adds that a full debate will still take place with the commission and the council; parliamentary officials say that most committee members favour reducing regulatory and administrative requirements for GMOs.

Another plank of the white paper is a plan to build an EU electronic highway. Funding is one obstacle. But the biggest will probably be resistance among member states to the liberalization of telecommunications monopolies, while Desama wants assurances that the highway will cater for remote areas and small companies.

Declan Butler

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The dispute on gene patents is still unresolved.

Greenpeace/Dave Adair

US health agencies at odds over trial deaths

Washington. Clinical trials last spring by the US National Institutes of Health (NIH) of the drug fialuridine, in which five patients died of liver failure, were carried out "with no noteworthy flaws", according to a panel of external scientists appointed by the (NIH) to investigate the tragedy.

The panel's report, submitted last week to Harold Varmus, the director of NIH, says that the studies were well designed, conducted and monitored, and that careful attention was paid to patient symptoms. "There were no errors in clinical judgement or the research process," the report says.

But the findings of the panel, which was jointly chaired by David Challenor, vice-president of the University of Florida, and David Kipnis of Washington University, St Louis, Missouri, conflict sharply with those of a parallel investigation by the Food and Drug Administration (FDA).

Last month, the FDA sent severely worded warning letters to Jay Hoofnagle, the principal investigator, and others in-

involved in the trials — including their sponsor, Eli Lilly of Indianapolis — alleging lax reporting of patients' symptoms and flawed procedures for gaining their informed consent (see *Nature* 369, 268;1994).

The five patients died suddenly during the third trial, which was abandoned. Subsequent investigation revealed that at least one patient on an earlier safety trial had died the same way, but the death had been attributed to the patient's illness, not to side-effects of the drug.

The NIH panel spent six months examining the reasons for the three NIH fialuridine studies, how well they were conducted, how the institutes responded to problems with the trials, and whether the five deaths from liver failure on the third study could have been avoided.

It found that there was "a justifiable scientific rationale" for all the studies, and that the research team's progression from the first study, carried out on HIV-positive patients, to the two later ones on patients

with hepatitis B, was "justifiable and appropriate at the time".

The fialuridine studies, it says, "represent the best of current practice in clinical investigations and exceeded regulatory requirements where such applied". The NIH response to the liver failures on the third trial "was entirely appropriate and swift, in accordance with the highest standards of patient care".

Critics of the trial have questioned the independence of the panel, which reported to the NIH director's advisory committee. "You can't trust self-investigations," says Sidney Wolfe, a former NIH researcher who works with the Washington pressure group Public Citizen. Congressman Ed Towns (Democrat, New York), chairman of the House of Representatives subcommittee on human resources and intergovernmental relations, called the report a "whitewash".

But members of the seven-strong panel — six of them senior professors from US medical schools — point out that their professional reputations were at stake. "We were all ready to go public with bristling criticism if it was warranted," says Richard Corlin, a Californian physician.

Kipnis says that Varmus had made it clear to the panel that its task was to find out the facts and report them. "Any suggestion of whitewash would be offensive to every individual on the panel."

NIH panel members characterize the FDA investigation process as primarily an audit of figures, charts and times which has, for example, pinpointed minor discrepancies in the reporting of symptoms, even though they were not material to the tragic outcome of the third trial.

But Jim O'Hara, a spokesman for the FDA, points out that the agency talked to the investigators and the sponsors, as well as looking at the data. "We issued compliance letters because we believe there to have been significant violations of our regulations, and we stand by our letters," says O'Hara.

The NIH panel acknowledged that the two studies came to "apparently different conclusions", but says the FDA ignored repeated requests for a copy of its unpublished audit of the trials. The FDA has given Hoofnagle and Eli Lilly until 29 June to respond in detail to its warning letters.

The high profile of the fialuridine case, and the appearance of various overblown accounts on television and in Congress, has done little to ease long-standing tensions between the two federal agencies. But the impasse may be broken in six months' time when the Institute of Medicine delivers its verdict on the fialuridine case to the cabinet member responsible for both agencies, Donna Shalala, the Secretary of Health.

Colin Macilwain

'Irreproducible' team clones a rival

San Francisco. The American staff of *The Journal of Irreproducible Results* (JIR), the oldest and best-known satirical science journal, have left to set up a competing magazine after several years of disagreements with its British publisher. The 40-member board of scientists, which includes seven Nobel prizewinners, went with them.

The first electronic issue of the new *Annals of Improbable Research* (AIR), including news of what it described as the "revolt of the mad scientists", appeared on the Internet last week. The first paper edition is expected in the autumn.

But Jon Conibear, deputy managing director of Blackwell Scientific Publications in Oxford, England, which publishes *JIR*, denies claims that the future of the original journal is uncertain beyond the next edition. "The journal will continue to be published as before," he says, claiming that the launch of *AIR* is not seen as a threat to *JIR*.

According to editor Mark Abrahams, the split is the result of long-lived antagonism with Blackwell. "It made them very uncomfortable that, of all the journals they published, this one was intentionally funny," says Abrahams, who claims that a largely volunteer staff was carrying out all marketing and publicity for the journal with no help from Blackwell.

"It finally seemed clear to us that there was no reasonable way we could either improve that situation or buy the rights to the name," says Abrahams.

"We all therefore decided to do the only reasonable thing: start a new magazine — one that has no legal connection to the magazine we left behind, on which we all worked so hard and loved so much."

But Conibear says that Abrahams was asked to leave as a result of irreconcilable editorial differences with *JIR*'s Chicago-based owner George Sherr. Ownership of *JIR* is due to revert back to Sherr after the June issue. Sherr is expected to announce a replacement editor later this month.

JIR was founded in 1955 by Alex Kohn, a virologist at Tel Aviv University, best known for his scientific paper proving that the North American continent is likely to sink under the accumulated weight of stockpiled National Geographic Society magazines. He eventually sold the rights to *JIR*, but remained as editor until 1989, when Abrahams took over. He and Abrahams are the co-founders of *AIR*.

At the time of the revolt, *JIR* had a circulation of 17,000, including the electronic issue. Abrahams claims that it was once as high as 40,000, although Conibear disputes this figure.

The new publication takes with it the infamous Ig Nobel Awards, given out each year to people whose achievements in science "cannot or should not be reproduced". The awards ceremony will henceforth be sponsored jointly by *AIR* and the Massachusetts Institute of Technology Museum, where the award ceremony takes place.

Joel Shurkin

Critic still lays blame for AIDS on lifestyle, not HIV

London. Haemophiliacs who have died of AIDS after being given blood infected with HIV have probably succumbed to "a whole lot of things that could be described as AIDS", rather than the virus itself, according to Kary Mullis, the Californian chemist who is backing the London *Sunday Times*' campaign against what it describes as the "HIV myth".

Speaking in London last week, Mullis also suggested that the deaths of HIV-positive children were in fact caused by their treatment with the drug AZT. And the epidemic of AIDS in Africa, which he disbelieved, was probably the result of misleading AIDS tests due to a cross-reaction with malaria antibodies; "most Africans have antibodies to malaria which look like antibodies to AIDS."

According to Mullis, the only genuine cases of AIDS have been those in the gay community in centres such as San Francisco and New York. The disease, he said, is probably the result of the breakdown of the immune system following exposure to a wide collection of retroviruses accumulated through multiple contacts in bath houses.

But the idea that HIV was the "probable cause of AIDS" was not a scientifically proven fact, he said. Nor, indeed, was there any clear evidence that AIDS was spread through sexual contact. "I think we [spread] retroviruses by our lungs, not by our genitals," said Mullis, who was awarded last year's Nobel prize for chemistry for his invention of the polymerase chain reaction (PCR) technique.

Mullis gave his views, which are eyed sceptically by most members of the scientific community, to a meeting organized by the newspaper entitled "Why there is still an HIV-AIDS controversy".

In common with others who claim that HIV has wrongly been identified as the "probable cause" of AIDS, such as the molecular biologist Peter Duesberg, he claimed supporters of this "hypothesis" are primarily

motivated by financial and career ambitions.

Mullis said his own scepticism stemmed from an inability to find a single widely-accepted scientific citation for the statement — which he himself used in papers describing the use of PCR for detecting HIV — that HIV is the probable cause of AIDS.

Modern science is based on the ability to demonstrate facts to an audience, however sceptical, he said. Any postgraduate student who had written a convincing paper demonstrating that HIV "causes" AIDS would be credited with having published "the paper of the century", he said.

Mullis's remarks received enthusiastic support from several members of the audience, who expressed their own criticism of the "establishment" handling of the AIDS crisis. But scientists in the audience remained deeply sceptical. Tim Oliver, a cancer physician, questioned Mullis's statement that HIV-positive haemophiliacs have died of AIDS for reasons unrelated to their exposure to the virus. He was unconvinced by Mullis' reply.

Many scientists had rejected invitation from the *Sunday Times* to attend the meeting. Robin Weiss, for example, director of research at the Institute of Cancer Research, wrote that Mullis "has about as much knowledge and authority to speak on the cause of AIDS as William Shockley, the Nobel Laureate who invented the transistor, had when he used to speak on the intellectual superiority of whites compared to blacks."

Mullis himself appeared unfazed by such criticisms. His main concern was that the efforts of "a lot of talented people in the medical profession had been diverted" by misconceptions about HIV. "It is not a scientific fact, or even a supposed fact, that HIV is the probable cause of AIDS," he said, unrepentant in the face of those who complained his ideas were undermining the activities of AIDS health-workers around the world.

David Dickson

Lay panel sought to give verdict on 'engineered' plants

London. A search was launched last week for 15 "very special 'ordinary' people", to make up a lay panel to give its views on behalf of the British public on the use of genetic engineering in plants. Panel members will be recruited through advertisements on local radio and in regional newspapers.

Britain's first 'consensus conference' is being funded by the Biotechnology and Biological Sciences Research Council (BBSRC) and organized by the Science Museum in London. It is based on procedures developed in Denmark, where they have been used by the Danish Board of Technology in fields such as food irradiation and mapping of the human genome.

Volunteers must have no vested interest in biotechnology, but must be interested in the social and ethical implications of its use. They must also be willing to give up at least three weekends to the project, for which they will receive only expenses.

During the first weekend, members of the panel will be given basic information about plant biotechnology, enabling them to identify areas of concern. At a second meeting, the panel will prepare written questions to be put to a pool of experts.

The process will end with a public conference at the beginning of November at which members of the panel will cross-examine the selected experts directly. The lay panel will then prepare a report summarizing its views.

"We feel that 15 panel members will provide sufficient scope to bring diversity of background and experience," says George Gaskell, senior lecturer in social psychology at the London School of Economics, who is responsible for the selection of members of the lay panel.

John Durant, assistant director of the Science Museum, and a member of the project's steering committee, said interest has already been expressed in a similar consensus conference looking at biotechnology in medicine and genetics.

The BBSRC has contracted the organization of the process to the Science Museum to avoid criticism that the project is an industry set-up. For the same reason, it has been keen to balance the members of the steering committee, which will include both a senior executive from Zeneca Seeds, and Harriet Kimble, senior deputy chairman of the Consumers Association.

According to Kimble, the procedure is intended to reduce the chances of new genetically engineering products being rejected by consumers by ensuring that they feel they have been properly consulted beforehand.

Maggie Verrall

New rules on CERN membership 'essential'

London. A change in the rules allowing physicists from countries that are not members of the European Laboratory for Particle Physics (CERN) in Geneva to participate in experiments at the laboratory is essential if international agreement is to be reached on the funding of the proposed Large Hadron Collider (LHC), according to a report published this week by the UK Parliamentary Office of Science and Technology.

At present, foreign physicists are able to work at CERN, even though contribut-

ing little to its cost, in exchange for access by Europeans to their own machines. But given the decline in state-of-the-art accelerators, continuing present arrangements is "not an option", says the report.

Keeping the status quo would mean that "Japan, America, Canada, China, India and other non-members would be heavily subsidized". The future success of CERN will depend on finding a way of encouraging non-member states to participate in CERN "on a more equitable basis", says the report. □

Japan eyes plans for robot-built Moon base

Tokyo. Determined not to be caught flat-footed if the international Space Station planned by the US National Aeronautics and Space Administration (NASA) finally gets the axe, the Japanese have made a counter-proposal. A report just published in Tokyo calls for Japan to construct a manned lunar base by the year 2024.

Construction would be carried out in three phases over 30 years, at a cost of around ¥2,900 (US\$30 billion). The main idea is that the Moon base would be built entirely by robots, exploiting Japan's advanced automation technology. Materials would be ferried to the site using 87 H-II rockets, Japan's indigenous launch vehicle, which had its successful maiden flight in February.

The first phase of the project, from 1999 to 2005, would land a lunar rover to survey possible landing sites. During the second phase, from 2006 to 2016, pilot plants would be constructed for the production of food, oxygen and energy. The final phase, from 2017 to 2023, would see the building of full-scale facilities, including living quarters for people.

The proposal includes a permanent staff

of six. International partners would be welcome, but — significantly — are not seen as an essential pre-condition, a reflection of Japan's frustration over its involvement in international projects in the recent past.

The proposal comes from the newly-formed Lunar and Planetary Society, a think-tank whose members include academics, government researchers and engineers from 20 of Japan's largest heavy engineering and construction companies. The society was established to follow up a large-scale study on lunar exploration and exploitation sponsored by the Institute for Future Technology, an independent non-profit organization, and carried out between 1989 and 1991.

The principal advocate of the proposal is Ryojiro Akiba, director-general of Japan's Institute of Space and Astronautical Science (ISAS). But according to Nobuki Kawashima, who represents ISAS on the steering committee of the Lunar and Planetary Society, Akiba is acting as an independent individual, not in his official capacity.

The reason for the distinction, explains Kawashima, is that "scientists at ISAS are still very cool" towards the plans for a lunar

base. "When we discuss lunar base construction, every time the question arises, what is it for?" he says.

Another question that is difficult to answer is whether the plan is feasible. Until now, the general consensus has been that the construction of a lunar base would cost at least ¥29 trillion. The Japanese report says that unmanned, robotic construction can be used to reduce the price by an order of magnitude, to ¥2.9 trillion.

But the plan is also based on the use of the H-II rocket, considered to be by far the most expensive launch vehicle in the world. "If we use the current H-II, it is inconceivable that we meet the proposed budget," Kawashima concedes.

Despite such discrepancies, the lunar base proposal is said to have been greeted favourably by members of the Space Activities Commission, the body that makes recommendations on space policy to the Japanese government's Science & Technology Agency. Japan reviews its space policy every five years, and it is up for review this year. A decision on whether officially to adopt the lunar base proposal is expected in the autumn.

Bob Johnstone

Sex discrimination among India's engineers confirmed

New Delhi. Women engineers in India face discrimination from the time they enter engineering colleges, and it continues throughout their careers, according to a government-sponsored study. Not only are they treated differently from their male counterparts during their training, but many industrial companies still treat them as misfits, only giving them jobs for which there are no male applicants.

The two-year study was sponsored by the Department of Science and Technology (DST), and carried out by Mrs P. P. Parikh and S. P. Suhatme, professors in the mechanical engineering department of the Indian Institute of Technology in Bombay. It urges women engineers to form a separate organization to defend their rights.

The need for the study originated in the rapid increase in women engineers in India over the past 15 years. There were fewer than 1,000 in 1975, but by 1990 the figure had risen to 19,180, and it is expected to reach 53,000 by 1996.

As numbers started to rise rapidly, personal and career problems began to surface, say the two authors. Having identified more than 16,000 women engineers, the authors sent out questionnaires to 5,000 of them, and received replies from 2,753.

These provided examples of the different types of discrimination in engineering colleges. For example, the study found that

female students are often denied the project of their choice by male teachers, and required to take summer courses in 'design training', as companies are reluctant to give them practical experience on the shop floor. In one college, the three women in a class of 70 were even excluded from industrial tours.

Problems escalate when woman engineering graduates start to look for jobs. Campus interviews are often restricted to

Kerala University eventually became an office clerk at US\$40 a month after she failed to get a job as an engineer after three interviews.

Partly for these reasons, unemployment among female engineers is 26 per cent — about five times higher than among males.

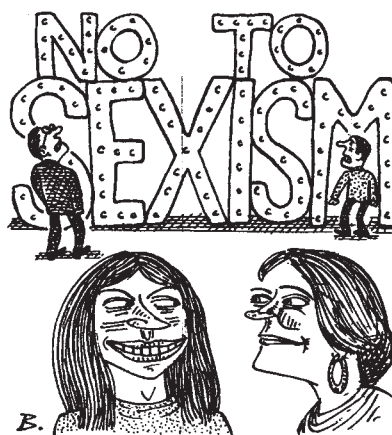
As a result of discrimination, many women engineers are forced to take teaching jobs, or to join the civil service. The problem is repeated in private industry, where women make up 20 per cent of the workforce, but nearly 60 per cent carry out secretarial or administrative duties.

Despite recent signs of improvement, many women engineers "feel they are not getting what they deserve in terms of promotions, salaries, as well as professional recognition and incentives", the report concludes.

It proposes two strategies to improve their status. The first is to increase the number of women enrolling in engineering colleges through 'awareness programmes' for teachers and parents and an increase in the number of places available for women.

The second would be to include women on interview boards, and to establish a minimum quota of women in engineering companies, if necessary enforced through legislation. The National Commission on Women has agreed to set up a separate body to ensure that these recommendations are implemented.

K. S. Jayaraman



men, and many companies openly declare that women need not apply.

If called for interviews, women face questions such as: "How will you manage both family and job?" A female graduate from

Clinical analysis in Egypt

SIR — Developing countries suffer from many health problems, principally because of the lack of infrastructure and old-fashioned regulations. The Egyptian journal *Al Ahram* published in its international edition of 28 February the health minister's promise to amend the law, restricting the clinical analysis profession to physicians only.

In the same article, the president of the medical college in Egypt added that only physicians are qualified to be clinical analysts. His first justification was longer graduate studies in medicine (one year more) than in pharmacy, but he took no account of the period of specialization of both. Second, he presented the misguided argument that in the developed countries only physicians are authorized to be clinical analysts. Such pronouncements show ignorance of the education systems in developed countries.

Law 367 (1954), which regulates this specialty in Egypt, recognizes only four specialties: medical chemistry, bacteriology, pathology and clinical pathology, and needs to be updated. Changes might include (1) recognition and legislation of recent technologies such as molecular biology, genetic cytology, immunology and so on; (2) improving the quality of health professionals by extending the period of specialization to 3 or 4 years (instead of 1 or 2 years) regardless of whether they are pharmacists or physicians; (3) improving the service quality of the laboratories by external quality control; (4) including parasitology as a specialty, in a country where parasitic infections constitute a profound health problem; and (5) repealing Law 415 (1954), which prohibits specialized pharmacists from taking samples including blood and urine or even stool samples from patients for analysis — according to this law, they must be taken by a physician.

Salah Abdalla
Mutualidad Mahonesa,
Mahon, Menorca,
Balears, Spain

Leprosy vaccine

SIR — According to Bloom *et al.*¹ the two leprosy vaccine strains, ICRC and M.w., showed identical bands in the RFLP analysis at only 2 or 3 points. But this is an understatement. The authors used two enzymes (*Pst*I and *Bst*EII) and two probes, 3.6 kb *Mycobacterium leprae* 65 kD antigen and 65 kD *M. tuberculosis* protein gene probe, and obtained identical bands in RFLP studies with reference to two strains in every experiment, under all four

sets of conditions². When the results are computed, it is clear that the RFLP fragments are identical at 7 bands. The authors have further studied RFLPs with two additional enzymes, *Eco*RI and *Bam*HI (data not shown), implying that the results are similar. In that case the RFLP identity will be seen, at least, at four additional points. Thus, in their experiments, the authors must have observed at least 11 identical bands. Not a single different RFLP band was seen between the two vaccine strains. Further, their pattern differed completely from all other mycobacteria studies in the paper. When these observations are taken in totality, the probability that any two unrelated organisms will give such a matching pattern is extremely low.

M. G. Deo
Cancer Research Institute,
Parel, Bombay-400012, India

1. Bloom, B. R., Jacobs, W. R. Jr & Clark-Curtiss, J. E. *Nature* **368**, 579 (1994).
2. Grosskinsky, C. M., Jacobs, W. R. Jr, Clark-Curtiss, J. E. & Bloom, B. R. *Infect. Immun.* **57**, 1535–1541, (1989).

Meteoric fall

SIR — Brown *et al.*'s letter and Hughes's commentary on the fall of the Peekskill meteorite (*Nature* **367**, 624 & 596; 1994) fail to mention one of the more fascinating aspects of the collision — namely the response of Ms Knapp's insurance company to the claim that her 1980 Chevrolet Malibu had been damaged by a 12-kg chondritic meteorite.

Incidentally, have our actuarial colleagues been forced to recalculate the odds of a meteorite colliding with a car or, indeed, a person as a result of this one-off incident? If they haven't, it may now be time for a small wager . . . or the setting up of a meteorite mutual protection insurance fund.

Mark Fowlson
Medical Editorial Unit,
Department of Health,
London SE1 6EF, UK

Gene safety

SIR — My reaction to the article "Error of judgement over gene safety rules" (*Nature* **367**, 49; 1994), was one of sympathy for the scientists involved. What the article did not say was that the researchers had used a replication-defective adenovirus vector expressing the SV40 large T protein. The activities of this protein have been worked on for more than 20 years and, although all the pathways for its mode of action are still unclear, a lot has been learnt. The SV40 virus expressing large T is semipermissive for humans and was in fact mistakenly inoculated into

humans in the late 1950s during the early poliovirus and adenovirus vaccine trials. The levels of inoculated SV40 were sufficient to produce an immune response, yet no adverse effects were observed.

One can in fact work safely, according to guidelines, with titres of 10^8 to 10^{10} pfu ml^{-1} of SV40 on the bench. The virus is unable to produce tumours in any mammals except the hamster, and then only with large doses.

It was stated in your report that Health and Safety Executive (HSE) officials acknowledge that there was a "theoretical" risk, but the department "failed to undertake a sufficient assessment of the risk to humans". It is surprising that the adenovirus recombinant, which is unable to replicate after infection of a cell and is therefore very unlikely to recombine with an endogenous adenovirus, was thought to have the level of risk to close a laboratory. I believe the risk of infection with the virus has been assessed in the best way possible, by the inoculation of individuals with infectious virus and the use of this virus by scientists for the past 35 years. The HSE obviously thought it was time to prove its usefulness as a controlling body, but picked a very weak test case, which will make it difficult to be taken seriously when justifiable situation arises.

Dennis J. McCance
Department of Microbiology
& Immunology, & of Oncology,
University of Rochester,
Rochester, New York 14642, USA.

Pedant mavens

SIR — I am a regular reader of *Nature* and of William Safire's "On Language" columns in the *New York Times Magazine*. I imagine that Safire, an acknowledged language maven, would be amused to find himself credited with the coinage of the word 'maven', as he was in Christopher Longuet-Higgins book review ("The talking ape", *Nature* **368**, 360; 1994).

While Safire is no doubt responsible for popularizing the phrase 'language maven', 'maven' is in fact Yiddish for an expert, connoisseur or judge, according to my *Modern English-Yiddish, Yiddish-English Dictionary* (Uriel Weinrich, Schocken Books, New York, 1977). In common usage 'maven' frequently connotes a self-appointed know-it-all, which is apparently the sense Stephen Pinker intended when he titled one of the chapters in his recent book (*The Language Instinct: How the Mind Creates Language*, Morrow, 1994) "The Language Mavens" and then went on to disparage the phrase as an oxymoron.

Sharon R. Seiler
Memorial Sloan-Kettering Cancer Center,
1275 York Avenue,
New York, New York 10021, USA

Cell-cycle regulation by numbers

A mathematical model of the cell cycle, of great interest in itself, may be a first step towards the much more ambitious models people will be building in the decades ahead.

ONE of the deepest pits into which journalists can fall is that of what may be called prescriptive conceit. That is the tendency of people who have been observers of some scene for many years to believe that they know better than the actors currently on stage what lines should next be said. The result is pompous and hortatory prose littered with verb-qualifiers such as 'should' and even 'must'. The tendency is a pitfall because experience also shows that the only people who read that kind of prose are those who recognize at the outset that they will agree with the prescriptions; other people's beliefs will be unaffected.

That is by way of half an apology (some may even consider it to be a negative one) for this part of *Nature's* campaign for a better regard among molecular biologists for the quantitative aspects of their exciting work. OK, these guys are indeed telling the rest of us how the cell works — which molecules do what to which other molecules. From time to time, there are even a few scraps of structural information that lend conviction to hand-waving accounts of 'mechanism'. But they are still naming the parts of the machine, without explaining to the rest of us why it works like that.

Specifically, there are two dangers in the current cult of the qualitative in molecular biology. One is that it gives the impression that everything is cut and dried. Replication of DNA is faithful to the template, the translation of mRNA into protein similarly reflects what is written in the genome (give or take a modicum of ambiguity in nuclear splicing) and the question of how external influences may modulate the genomic autonomy of the cell must await the naming of the parts concerned.

The second danger is more serious: the practitioners are at risk of missing important truths about the systems they are studying. Two years ago, an article in this series singled out cell division as a phenomenon likely to be determined by 'quantitative triggers', not qualitative ones (*Nature* 355, 201; 1992). The same piece concluded with a plea that molecular biologists should resurrect the Law of Mass Action.

Independently of these urgings from the sidelines, Bela Novak and John J. Tyson from the Virginia Polytechnic Institute at Blacksburg now appear to have done just that (*J. Cell Science* 106, 1153–1168; 1993). (Novak really belongs to the Technical University of Budapest.) What they have done is to construct a mathematical model of the regulation of the cell cycle by the

biochemicals now known to be involved.

The still emerging tale of how the cell cycle is regulated is one of the most enthralling of the past few years, involving the interplay between experiments with amphibian oocytes and studies of the genetics of yeast strains defective in some aspect of cell-cycle regulation.

In essence, nevertheless, the story is straightforward. (The third edition of *The Molecular Biology of the Cell*, published earlier this year, has an excellent up-to-date account.) The key event in the life-history of a cell is its division into two replicas of itself, accomplished during 'M-phase' (where 'M' stands for mitosis). The rest of the cell cycle, called 'interphase', is from one point of view merely a preparation for mitosis, but cells such as neurons spend their whole time in that condition. Otherwise, interphase is when the cell's complement of DNA is replicated and the organelles required to send two daughter cells successfully on their way are manufactured.

What controls the oscillation of cells between interphase and mitosis? What else but a 'factor'? That was originally called MPF (for maturation promoting factor), but has now been identified as a dimer of two distinct protein molecules, a cyclin and a protein kinase related to that originally identified in fission yeast and known as cdc2. The activity of MPF as a kinase (or phosphorylating enzyme) depends on the manner in which its cdc2 component is itself phosphorylated, but some of the enzymes controlling that process have also now been identified.

Novak and Tyson are above all concerned that their model should allow for the two feedback loops recognized in experiments. First, active MPF stimulates its own production, which is positive feedback. But active MPF also stimulates the destruction of cyclin (by what is called 'ubiquitin-conjugated enzyme'), which is a negative feedback.

The result, inevitably, is a set of linear differential equations (in which the only differential coefficients are the rates of change of concentrations with time). In reality, there are 13 of them, but only 9 are independent (because of the assumption that the total concentration of cdc2 is constant throughout the cycle). Altogether, there are 18 rate constants for the chemical reactions involved, while it is possible to look for solutions of the equations only by saying something about the initial values of the 13 concentrations of the materials involved in

regulating the cell cycle.

The equations can be solved only numerically, but that is no great handicap. Moreover, the agreement with experiment seems remarkably good. For example, it is possible to extract from the equations a curve describing the conditions for equilibrium between total (bound or otherwise) cyclin concentration and that of active MPF. And that turns out to be unstable. There is some point at which a small increase of cyclin concentrations brings about a jump in active MPF concentration of nearly an order of magnitude — just the kind of trigger cell-cycle regulation needs.

Perhaps the most remarkable illustration of the fidelity of the equations to real life is their use in the simulation of the oscillation of the cell-cycle regulators in cell-free extracts from *Xenopus* eggs. That is the most remarkable phenomenon in itself. Like other amphibia, *Xenopus* embryos must grow into tadpoles autonomously, with no external support, so that mature oocytes must contain the biochemical machinery required for many successive cell divisions. So even if the nuclei are removed, so that there is no DNA to replicate, cytoplasm on its own will go through the motions of cell division. What Novak and Tyson find is precisely what they expected — an oscillation of their system with a period of about 80 minutes.

Where this will lead is easily imagined. Novak and Tyson's parameters have been chosen to match the *Xenopus* case. No doubt they are busily applying the equations to other systems. It will be interesting to learn how the rate constants vary from one system to another, but careful matching of the numerical results with experimental data should also help to identify the many points at which understanding is incomplete.

And that is only a beginning. The coupling of this model to the external environment of the cell is at present only rudimentary, for example, but putting that right will entail the addition of further equations. Indeed, some model of this kind is likely eventually to be the core of a mathematical model of the cell as such. Eventually, that will be a project exceeding in complexity the Human Genome Projects in their various forms, with some thousands of coupled equations. But cell biologists should not be disheartened by the prospect. Those who model the chemistry of the Earth's atmosphere or even that of interstellar molecular clouds are well used to systems that are at once as mathematically simple and almost as complex.

John Maddox

A clasped embrace

Dinshaw J. Patel

HUMAN chorionic gonadotropin (hCG) is a placental glycoprotein hormone that induces the secretion of progesterone to sustain the early weeks of pregnancy¹. But despite its central role in reproductive physiology, we have had little definitive information about the global folding of

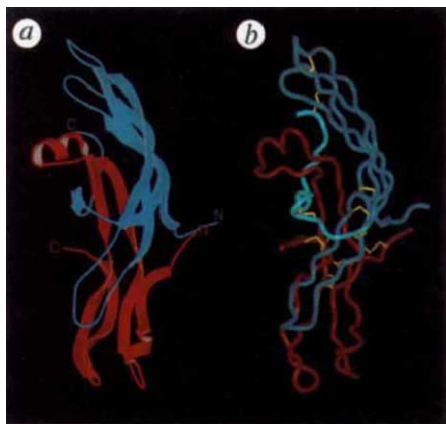


FIG. 1a, Ribbon diagram of the hCG heterodimer with α -chain (5–89) in red and β -chain (2–111) in blue. The amino and carboxy termini of the chains are labelled and the disulphide bridges are omitted. b, The $C\alpha$ backbone trace of the hCG heterodimer viewed in a similar orientation to a. The α -chain is in red, the β -chain in blue; residues β (90–111) are in light blue and the disulphide bridges are in yellow. (All figures courtesy of Hao Wu).

hCG, the interactions between its α - and β -subunits, the elements of its structure that interact with its receptor, and the molecular events that lead to signal transduction.

This barren landscape is now immeasurably brightened by the appearance of two reports of the X-ray crystal structure of hCG, one on page 455 of this issue², the other in *Structure*³. They go a long way towards linking structure and biological function for the family of glycoprotein hormones that stimulate ovarian and testicular function. The folding topology of hCG is most unusual in that it has unexpected folding motifs, interlocking subunit–subunit interactions and symmetry relationships, as well as a surprisingly high ratio of protein surface to hydrophobic core. Not least, the newly solved structure will open up fresh avenues for structure-based approaches to the design of drugs for both contraception and the stimulation of fertility.

Human chorionic gonadotropin is a noncovalent heterodimer composed of a 92-amino-acid α -subunit (with five disulphide bridges and two *N*-linked glycosylation sites) and a 145-amino-acid β -subunit (with six disulphide bridges and

two *N*-linked and four *O*-linked glycosylation sites)¹. Carbohydrates account for 30–35 per cent of the total mass of hCG, and crystals^{4,5} have been successfully grown following hydrogen fluoride treatment that results in a roughly 50 per cent reduction of the carbohydrate content of the hormone⁶. Laphorn *et al.*² have solved the hCG structure at 3.0 Å resolution, using multiple isomorphous replacement combined with maximum entropy solvent flattening techniques². A less conventional approach was taken by Wu *et al.*³, who produced the selenomethionyl protein in mammalian cells, crystallized it under anaerobic conditions and applied multi-wavelength anomalous diffraction measurements to generate the hCG structure at 2.6 Å resolution. This higher resolution structure also allowed the identification of 64 solvent molecules. The good news is that the same global fold has emerged from these two independent investigations; both groups have been able to trace residues 5–89 in the α -subunit and residues 2–111 in the β -subunit, and have defined the intermolecular interactions that stabilize the hCG $\alpha\beta$ heterodimer.

The heterodimer adopts an elongated shape consisting primarily of β -sheets with the two subunits intertwined into each other (Fig. 1). The core structure of each subunit contains a cystine knot motif^{7,8}, where a disulphide bridge connecting two strands penetrates into an eight-residue circle generated by a pair of disulphide bridges connecting two other strands. Extending out from the cystine knot in the α - and β -subunits are a pair of highly twisted hairpin loops, in almost parallel planes, which project in one direction, and a long, flat hairpin loop which projects in the opposite direction. The unexpected similarity of the global folds of the α - and β -subunits is best visualized by superpositioning the subunits as shown in Fig. 2a. The subunit superposition is

striking in the cystine knot region (Fig. 2b), whereas differences are observed in the size, tip dimensions and pairing alignments within related hairpin loops. The long loop in the β -chain (blue, Fig. 2a) is much more open than its counterpart in the α -chain (red, Fig. 2a), with the latter having a sharp turn in the β -sheet at a proline-rich segment.

The reciprocal alignment of α - and β -subunits at the dimer interface generates a quasi-dyad symmetry axis (Fig. 1). The subunits are intimately associated over their interacting surfaces, and in the process bury a surface area of some 4,000 Å². The intermolecular interactions are of the β -sheet type at the heart of the interface, as well as in a segment where the disulphide-tethered carboxy-terminal tail

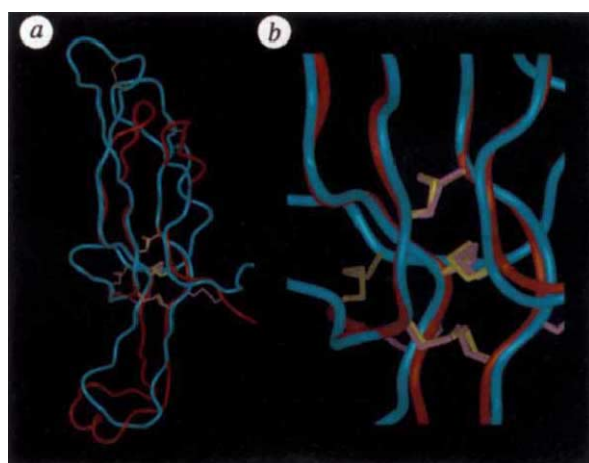


FIG. 2 Superposition of individual subunits. a, View of the α -subunit (in red, with magenta disulphide bridges) and β -subunit (blue with yellow disulphide bridges) of hCG following superposition of $C\alpha$ positions in the central cystine knot of both subunits. b, Enlargement of a showing the common folds that are adopted by the cystine knot segments of the α - and β -subunits.

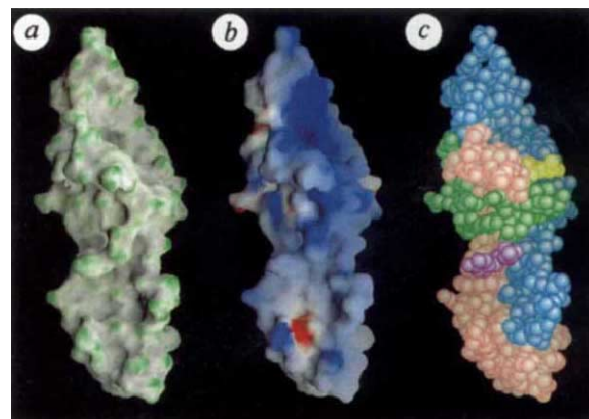


FIG. 3 The receptor-binding face of the hCG heterodimer. a, View showing the surface curvature with the convex, concave and planar surfaces in green, grey and white respectively. b, View showing the surface electrostatic potential, the most positive regions being blue and the most negative regions red. c, Space-filling view of the heterodimer (α -chain in pink, β -chain in blue), emphasizing the residues involved in receptor binding (α (88–89) in magenta, β (94–111) in green) and in signal transduction (Asn 52 and its attached carbohydrate in yellow).

of the β -subunit (light blue in Fig. 1b) embraces a long loop of the α -subunit. This 'seat-belt' arrangement has implications for possible *in vivo* folding pathways for hCG $\alpha\beta$ heterodimer formation, because this process will be governed by the precise order of disulphide bridge formation. Another striking feature of the hCG heterodimer structure is the large surface-to-volume ratio, which results in a one-layer structure with a minimal hydrophobic core at the dimer interface.

The cystine knot motif and the loops emanating from it, as well as the seat-belt arrangement which clasps segments of the α - and β -subunits, should turn out to be characteristics of other members of the glycoprotein family of hormones. These include luteinizing hormone, follicle-stimulating hormone and thyroid-stimulating hormone, all of which are secreted by the pituitary. These hormones have a common α -subunit and homologous but distinct β -subunits, hCG being most closely related to luteinizing hormone. The hCG structure should provide insights into potential $\alpha\beta$ heterodimer alignments in all three relatives.

Both hCG and luteinizing hormone have the same G-protein-coupled receptor, with distinct sites on the hormones being implicated in binding and signalling functions⁹. The receptor has a long extracellular domain extending out from a seven-helix membrane-spanning segment; the coupled G-protein is involved in activation of adenyl cyclase to produce progesterone. A view of the surface curvature for the receptor-binding face of hCG is shown in Fig. 3a. The electrostatic surface potential of this face is markedly positive (blue in Fig. 3b), and should complement the negative potential associated with the acidic residues on the extracellular domain of the receptor. Further, residues α (88–89) (magenta, Fig. 3c) and β (94–111) (green, Fig. 3c) that have been implicated in receptor binding^{10–12} are adjacent to one another on this face of hCG.

Equally interesting is the adjacent location on the same face of the N-linked Asn 52 glycosylation site which may be involved in signal transduction¹³. This Asn

52–(GlnNAc)₂ site, which is located in a depression on the hCG surface, is in yellow in Fig. 3c.

The determination of the hCG structure is a landmark among studies of the architecture and interactions of glycoprotein hormones involved in regulating reproductive physiology. The intertwined nature of the $\alpha\beta$ heterodimer interface and its minimal hydrophobic core, coupled with the unprecedented seat-belt wrapping of a tethered segment of one subunit around a segment of its partner in the heterodimer, opens up new challenges for those investigating protein folding. The central involvement of hCG in the early stages of pregnancy makes it a key molecule whose function can be mod-

ulated through selective binding of agonists to stimulate fertilization or of antagonists to induce contraception. Structure-based drug design strategies can now be initiated to address these goals.

Finally, the stage is set for attempts to crystallize and solve the structure of the complex of hCG and its extracellular binding domain on the receptor. That research agenda could eventually provide insights into hormone–receptor interactions on the pathway to signal transduction. □

Dinshaw J. Patel is in the Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA.

SOLVATION DYNAMICS

Wet chemistry

James T. Hynes

WHEN a chemical reaction occurs in solution, the transition state barrier must be crossed *en route* from reactants to products. Chemists have long known that the height of this barrier, which exponentially affects the reaction rate, can be extraordinarily dependent on the solvent, particularly for reactions in a polar solvent in which there is charge displacement and thus strong electrostatic interaction between the reactants and the solvent molecules¹. But solvation dynamics should also be important. Water, the commonest of solvents, is a case in point: how fast can its molecules reshuffle to suit the changing charge of the reaction intermediates? On page 471 of this issue, Jimenez *et al.*² show experimentally that the first, inertial response — tilting or twisting of the solvent molecules, rather than diffusive, reorientational motions — dominates aqueous solvation dynamics.

The exploration of solvation dynamics is of fairly recent vintage. For the past decade and a half, experimental and theoretical chemists have been trying to go beyond the standard notion of 'equilibrium solvation', instead probing just what the molecules are doing as the reaction progresses through the transition state and how this influences the rate.

To put this in context, consider how the well-known transition state theory¹ views a solution reaction, illustrated here by the bimolecular substitution reaction $\text{Cl}^- + \text{CH}_3\text{Cl} \rightarrow \text{ClCH}_3 + \text{Cl}^-$ in water^{3,4}. In effect, it is supposed that as the new bond is formed and the old one is broken, and especially as the electrical charge is internally redistributed, the surrounding solvent molecules are always equilibrated to the reaction solute system $[\text{ClCH}_3\text{Cl}]^-$ as transit through the transition state occurs — this is the equilibrium

solvation assumption mentioned above. Equivalently stated, it is assumed that the water solvent dynamics are fast compared with the (very brief) time spent by $[\text{ClCH}_3\text{Cl}]^-$ in the critical barrier region. If this condition is violated (and it is⁴), then the rate constant depends not only on the barrier height but also on the non-equilibrium dynamics of the solvent. And the actual rate constant will be less than transition state theory would predict. This can be understood in terms of the 'friction' that the reaction system experiences because of the non-equilibrium state of the solvent in the barrier passage; the larger that friction, the greater is the departure from the conventional theoretical rate^{5,6}.

Just how fast can a solvent equilibrate to a changing charge distribution of a solute within it? A powerful experimental avenue to answer this question is the study of time-dependence fluorescence (TDF) by ultrafast laser spectroscopy^{7,8}. Typically, a dye molecule in solution is initially electronically promoted to an excited state whose charge distribution is different from that of the ground state. But now the solvent molecules are out of equilibrium with the new charge distribution, and must begin their dynamic readjustment to reach equilibrium. When the solute's excited state is fluorescent, this process can be followed by the time-dependent shift of the fluorescence frequency, the dynamic Stokes shift. In a real sense, this probe is the dynamic — and hence more complex — generalization of the various static spectroscopic probes which have long been exploited to characterize solvent polarity directly¹.

Early theoretical models of TDF were based on simple, nonmolecular, dielectric continuum descriptions of the solvent,

- Pierce, J. G. & Parsons, T. F. *A. Rev. Biochem.* **50**, 465–495 (1981).
- Lapthorn, A. J. *et al. Nature* **369**, 455–461 (1994).
- Wu, H. *et al. Structure* **2**, 545–558 (1994).
- Harris, D. C. *et al. J. Biol. Chem.* **264**, 6705–6706 (1989).
- Lustbader, J. W. *et al. Biochemistry* **28**, 9239–9243 (1989).
- Manjuneth, P. & Sairam, M. R. *J. Biol. Chem.* **251**, 7109–7115 (1982).
- Murray-Rust, J. *et al. Structure* **1**, 153–159 (1993).
- McDonald, N. Q. & Hendrickson, W. A. *Cell* **73**, 421–424 (1993).
- McFarland, K. C. *et al. Science* **245**, 494–499 (1989).
- Chen, F., Wang, Y. & Puett, D. *Molec. Endocrinol.* **6**, 914–919 (1992).
- Campbell, R. K., Dean-Emig, D. M. & Moyle, W. R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 760–764 (1991).
- Moyle, W. R. *et al. Nature* **368**, 251–255 (1994).
- Matzuk, M. M., Keene, J. L. & Boime, I. J. *J. Biol. Chem.* **264**, 2409–2414 (1989).

and predicted an exponential time approach of the Stokes shift to its final equilibrium value. (Long-chain alcoholic solvents have their own special features⁹, so I'll exclude them from the present discussion.) In turn, most TDF experiments have been interpreted as indeed showing exponential behaviour, with time constants roughly in accord with the dielectric continuum picture.

But molecular dynamics computer simulations of TDF in model solute-solvent systems have shown a quite different picture. In particular, the observed solvent dynamical behaviour was bimodal rather than exponential, with a very significant rapid initial component (from tens to hundreds of femtoseconds depending on the solvent), followed by a weaker and slower time development. Theoretical arguments^{10,11} have suggested that the initial component is 'inertial' and molecular in character (as explained below), whereas the longer-lived contribution comes from collective, diffusive motions as envisaged in dielectric continuum models.

To a solvation chemist, the term 'inertial' has come to signify two different sorts of motion. The first is the free streaming of the solvent molecules; in other words, the dynamics they would have even in the absence of intermolecular forces. Even these motions change the solvent-solute interaction, and thus influence the Stokes shift. The second type — which will be especially important for hydrogen-bonded solvents such as water and methanol — includes quasi-oscillatory motions such as librations, which are torsional oscillations in the field of the solute and the other solvent molecules. For water,

these will be rapid because the hydrogens executing the motion are so light.

So much for the calculations. But where was this important fast inertial component in the TDF experiments for water, acetonitrile and other solvents^{12,13}? Evidently, time resolution and calibration difficulties have hampered their observation. With the work of Jimenez *et al.*² for water and others for different solvents^{14,15}, these difficulties have at last been surmounted. Experiment, theory and simulation alike now generally agree that the fast component is there.

What then is the significance of such short-time dynamics for reactions in solution? For charge-transfer reactions with a moderate to high activation barrier and involving bond-breaking and/or bond-making, such as the example above, it should be precisely these rapid rearrangements that determine the friction responsible for deviations from standard transition state theory. The slower solvent relaxations cannot contribute on the short

timescale in which the fate of an attempted transition state passage is decided¹. For high-barrier outer-sphere electron transfers¹⁶, and where no bonds are made or broken, the situation is different in detail, but, as Jimenez *et al.* point out, it is still the short-time solvation dynamics that will be crucial for the friction and thus how much the electron transfer rate deviates from its (Marcus) transition state theory prediction¹⁶.

For proper perspective, however, I should stress that reaction rates can also be sensitive to the longer-time solvation dynamics probed by TDF. This will occur, for example, when electron transfer barriers are low, and passage through a transition state region occurs at a sedate enough pace for those dynamics to play a role^{17,18}. □

James T. Hynes is in the Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA.

RNA

Life after transcription

Kimberly Carr

DNA makes RNA makes protein. Of these three players, it was the first and last that were for some time taken to be the glamorous ones. But research into the starring role of RNA in regulation of gene expression is gathering momentum, and the vigour of the field was amply demonstrated at a recent meeting that was held in Aruba*.

The phenotype of a cell is determined by the proteins it makes. The first step at which these are selected is the decision of whether or not to transcribe a gene. But control doesn't end there, and the transcribed precursor messenger RNA has a long path ahead of it before a protein is made. Post-transcriptional regulation of gene expression encompasses many processes, from mRNA processing and decay to the initiation and elongation of translation, as well as protein modification and targeting. RNA and its structure lie at the centre of all these events.

The process begins when the transcribed pre-mRNA is spliced to produce mature mRNA. Screens of yeast mutants defective in splicing have revealed proteins involved in the assembly, stability and function of components of spliceosomes, as well as a protein involved in nuclear export of mRNA.

Once the spliced message has been exported to the cytoplasm, another step in regulation takes place — control of mRNA stability. Destabilizing sequences in the 3' untranslated regions (3'UTRs) induce rapid decay of mRNAs. These AU-rich elements (AREs) are typical of genes whose expression must be tightly regulated (such as the early response genes *c-fos*, *c-myc* and *c-jun*). The exact sequences mediating destabilization are unclear, although certain ARE-binding proteins seem to be involved in mRNA degradation. In other cases, RNA-binding proteins block degradation; for example, iron regulation of transferrin receptor (TfR) expression involves stabilization of TfR mRNA by the binding of an iron regulatory protein to specific elements in the 3' UTR.

A major pathway of mRNA decay in yeast begins with truncation of the poly(A) tail, followed by decapping and exonuclease digestion. A special sort of RNA decay, however, induced by nonsense mutations, seems to bypass the need for poly(A) shortening. The fact that this type of decay requires not only the nonsense codon but also a downstream sequence which may be a translation reinitiation site, points to a link between mRNA turnover and translation. Interestingly, although mRNA decay usually takes place in the cytoplasm, nonsense-mediated decay in mammalian cells apparently occurs in the nucleus after splicing, possibly at the time of RNA

1. Reichardt, C. *Solvents and Solvent Effects in Organic Chemistry* (VCH, Weinheim, 1988).
2. Jimenez, R., Fleming, G. R., Kumar, P. V. & Maroncelli, M. *Nature* **369**, 471–473 (1994).
3. Chandrasekhar, J., Smith, S. F. & Jorgensen, W. L. *J. Am. chem. Soc.* **107**, 154–163 (1984).
4. Bergsma, J. P., Gertner, B. J., Wilson, K. R. & Hynes, J. T. *J. chem. Phys.* **86**, 1356–1376 (1987).
5. Hynes, J. T. in *The Theory of Chemical Reaction Dynamics* (ed. Baer, M.) 171–234 (CRC, Boca Raton, Florida, 1985).
6. Whitnell, R. M. & Wilson, K. R. *Adv. comput. Chem.* **4**, 67 (1993).
7. Barbara, P. F. & Jarzeba, W. *Adv. Photochem.* **15**, 1–68 (1990).
8. Maroncelli, M. *J. molec. Liq.* **57**, 1–37 (1993).
9. Su, S.-G. & Simon, J. D. *J. chem. Phys.* **89**, 908–919 (1988).
10. Carter, E. A. & Hynes, J. T. *J. chem. Phys.* **94**, 5961–5979 (1991).
11. Maroncelli, M. *J. chem. Phys.* **94**, 2084–2103 (1991).
12. Castner, E. W., Maroncelli, M. & Fleming, G. R. *J. chem. Phys.* **86**, 1090–1097 (1987).
13. Kahlow, M. A., Jarzeba, W., Kang, T. J. & Barbara, P. F. *J. chem. Phys.* **90**, 151–158 (1989).
14. Rosenthal, S. J., Xie, X., Du, M. & Fleming, G. R. *J. chem. Phys.* **95**, 4715–4718 (1991).
15. Rosenthal, S. J., Jimenez, R., Fleming, G. R., Kumar, P. V. & Maroncelli, M. *J. molec. Liq.* (in the press).
16. Weaver, M. J. *J. molec. Liq.* (in the press).
17. Fonseca, T. & Ladanyi, B. M. in *Ultrafast Reaction Dynamics and Solvent Effects* (eds Gauduel, Y. & Rossky, P. J.) 380–392 (American Institute of Physics, New York, 1994).
18. Barbara, P. F., Walker, G. C. & Smith, T. P. *Science* **256**, 975–981 (1992).

*Post-transcriptional Control of Gene Expression: The Central Role of RNA Structure, Aruba, The Dutch Antilles, 29 April–3 May 1994. Details of the speakers at the meeting and their presentations can be obtained from John McCarthy, the workshop director, at the Department of Gene Expression, GBF, Mascheroder Weg 1, D-38124 Braunschweig, Germany. Phone: 49-531-6181430; fax 49-531-6181458.

export. In combination with the short-cut to decay, this would prevent production of an incorrect protein.

Prokaryotes have a great deal to teach us about mRNA stability. The structure at the 5' ends of prokaryotic mRNAs is clearly important for message stability and its regulation by specific endonucleases, although the precise mechanisms remain unknown. Some prokaryotic ribonucleases seem to work in complexes with other proteins — for example RNaseE is found with polynucleotide phosphorylase and with the chaperone protein GroEL. Work with human cells has uncovered activities that may be comparable to bacterial RNases.

Once a correctly processed and (at least temporarily) stable message is available, it must still be selected for translation. Detailed studies of eukaryotic translational initiation factors (eIFs) have allowed analysis of the RNA-protein interactions that guide the ribosome to the start of a reading frame. New RNA-binding proteins that act in translational control have been discovered. One of these, the p50 component of messenger ribonucleoprotein, may be a general repressor of translation, as well as a transcription factor; and another, previously identified as the La autoantigen recognized by autoantibodies from systemic lupus patients, might be involved in translational initiation at internal ribosome entry sites. Although the translation of most mRNAs depends on the mRNA cap and ribosome scanning from the 5' end of the RNA, internal initiation is the mechanism of choice for picornaviruses. The RNA sequences and protein factors responsible for internal ribosome entry in picornaviruses are now being identified. They may also be involved in translation of some mammalian mRNAs.

In prokaryotes, transcription and translation occur in the same compartment and are tightly coupled. Decoupling the two processes by using a very fast RNA polymerase leads to destabilization of mRNAs, suggesting that coupling of the two processes prevents such decay. In eukaryotes, where transcription and translation take place in different cell compartments, it is still unclear to what extent some form of coupling over the nuclear membrane is possible.

Eukaryotic translation can be down-regulated by phosphorylation of factor eIF-2 α , as occurs in response to stress, or of the elongation factor eEF-2, phosphorylation of which seems to be involved in insulin regulation of gene expression. But translation is enhanced during *Xenopus* oocyte maturation by protein interactions with specific signals in the 3' UTRs of certain mRNAs; these signals mediate selective polyadenylation and thus translation of the messages. Moreover RNA-protein interactions that inhibit transla-

OPTICS

Reflections on a summer's night

ALISTAIR B. Fraser of the Pennsylvania State University evidently has a fatigue-defying appetite for meteorological curiosities. The glistening tree shown here is a fine example of a phenomenon that he first spotted in the small hours of a summer's morning, whilst driving along a forested road in British Columbia. Fraser describes his experience in an article to appear shortly in a special issue of *Applied Optics* (vol. 33, 20 July 1994): in his words, he noticed that in the car's headlights "the trees of the normally Stygian forest began to glow as if snow-covered", a startling observation on a warm August night. Evidently some form of retroreflection from dew-covered leaves, the effect seemed oddly species-specific, even on a first inspection. Later nocturnal expeditions with a powerful flashlight (a proceeding that aroused dark suspicions in at least one local gamekeeper) showed that it favoured only certain types of conifer and a few shrubs such as the yew and rhododendron. The explanation lies in the contact angle of droplets on the leaves: as this rises above 90 degrees or so, the proportion of light from the car's headlamps that is reflected back towards the occupant increases, and for angles above about 140 degrees, the retroreflection becomes spectacular. Blue spruces show the glow particularly well, because the tangled rod-like morphology of the epicuticular wax responsible for their blue bloom also increases the contact angle. Fraser dubs the glow sylvanshine, in deference to the *heilighenschein* "familiar to those who wander grassy fields in the early morning". But whereas the sylvanshine on a tree is visible to anyone in the car or holding a torch, *heilighenschein*, the result of sunlight focused by dewdrops held above a leaf's surface by fine hairs, appears to each observer as a halo around the head of just his own shadow on the grass. In his *Memoirs* of 1562, Benvenuto Cellini, no expert on the laws of optics, interpreted this instead as a sign of divine grace towards himself; and, says Fraser wryly, evidently none of those to whom Cellini vouchsafed his secret dared mention the halo about his own head.



Lindsay Matthews

tion can be used to characterize and clone new RNA-binding proteins.

Our picture of prokaryotic translation is also changing, as the third release factor for translational termination (RF3), which stimulates the other two, has now been found. Other prokaryotic insights include translational 'jumping', by which the ribosome bypasses a 55-base segment of mRNA, the incorporation of selenocysteine into proteins, which also occurs in eukaryotic cells, and the translational auto-regulation of threonyl-transfer RNA synthetases by 'molecular mimicry'; here, both the protein and its cognate tRNA bind to a specific codon in the synthetase mRNA leader. Translation of some prokaryotic RNAs is inhibited by ribosome trapping, for which specific RNA pseudoknot structures are required.

In all of these processes, RNA structure is crucial. As studies of RNA crosslinking

and ribosomal protein protection improve our three-dimensional view of ribosomal RNA structure, an increasing number of correlations between structure and function are emerging. The importance of RNA structure in controlling processes as diverse as translational frameshifting, ribosome binding and antisense-mediated control of plasmid replication emphasizes the versatility of RNA.

RNA was once regarded as a mere passive product of the real controller of gene expression, transcription. But it is now clear that it is central to the actions of the spliceosome and ribosome, as well as all the other processes of post-transcriptional regulation. We can expect to see a great deal of action as scrutiny of RNA versatility becomes more intense. □

Kimberly Carr is an assistant editor at Nature.

Cascading electrons

Raphael Tsu

CONVENTIONAL semiconductor lasers are limited by their chemical make-up; they emit radiation at an energy roughly corresponding to the difference between the conduction and valence bands, so whole realms of the mid- to far-infrared are out of reach. Considerable ingenuity has gone into developing new semiconducting alloys to reach new wavelengths. But a team at AT&T Bell Laboratories, led by Federico Capasso, has now produced a working prototype of a 'quantum cascade' laser which may circumvent the problem¹. Here, the energy levels are defined by the widths of a sequence of quantum wells; and as the name suggests, electrons cascade from level to level of the resulting energy staircase.

The cascade laser is a III-V semiconductor heterostructure built of fine layers of gallium indium arsenide and aluminium indium arsenide by molecular beam epitaxy. It is a complex structure, with repeated 'cells' of three coupled quantum wells sandwiched between electron-injecting (and collecting) regions. The coupled wells can loosely be thought of as artificial molecules.

The figure shows schematically how the energy diagram might look in one such active region, where barriers of AlInAs alternate with wells formed by GaInAs. Electrons are injected by tunnelling into the first and narrowest GaInAs region, which sets the shortest electron wavelength and thus the highest energy (level 3). They can then relax into the lower state 2 of the adjacent well — this is the lasing transition. To maintain the population inversion between 3 and 2, electrons are siphoned off into the third quantum well with an energy level 1 just below that of 2. From here, they can tunnel into the broad, compositionally

graded region and head for the next active region to the right. So the system is essentially a four-level laser.

All of this takes some clever structural design. To achieve population inversion between levels 3 and 2, it is important that the energy difference between the states is higher than the optical phonon energy, so that optical phonon scattering processes are rare and the relaxation time is fairly long. (Transitions involving emission of an optical phonon are not ruled out, but if the phonon energy E_{ph} is much less than $E_3 - E_2$, they do involve a large change of momentum Δk , and the transition probability falls as $1/\Delta k^2$.) In contrast, $E_2 - E_1$ is designed to be close to E_{ph} so that electrons are efficiently removed from 2. Electrons in state E_1 should also have a short lifetime because of the many available states in the adjacent graded alloy region.

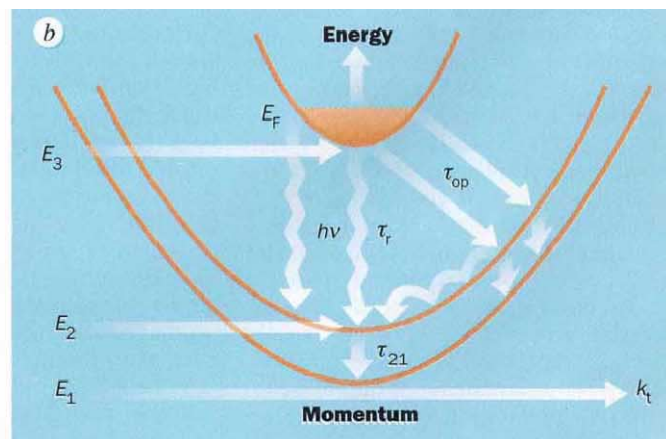
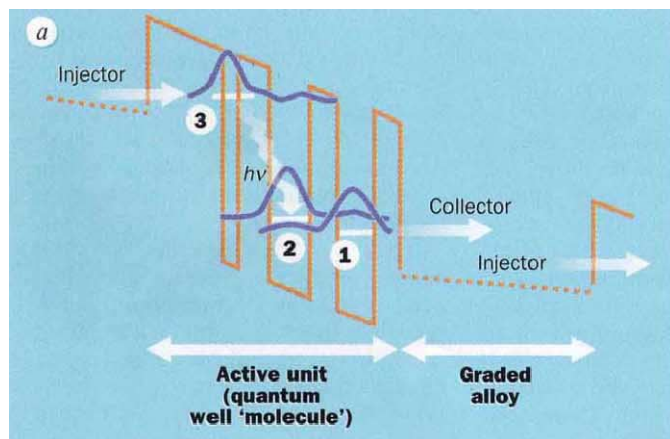
The contact serves both as a collector from the left-hand cell and an injector for the next cell to the right. But far from being a humble connector, it has a finely graded composition with a band gap that increases from left to right between the active regions. At zero bias, its band diagram has a sawtooth shape², but when a suitable voltage is applied across the whole structure, each of the collector-emitter regions becomes nearly flat and the whole thing resembles a staircase. Without this feature, electrons collected would be accelerated to the next cell, smearing the Fermi function and greatly reducing the efficiency of the resonant tunnelling injection.

Overall, the structure consists of 25 periods of coupled quantum-well cells. It can deliver 8 mW at a wavelength of 4.2 micrometres, so far only in pulsed operation and at low temperature (90 K is the

limit mentioned in the paper, although later unpublished results by the group reach a peak optical power of 30 mW at 102 K). As the radiative transition time between 3 and 2 is several orders of magnitude longer than the optical phonon relaxation time, population inversion is primarily maintained by high current injection, and the laser is a relatively inefficient device with substantial heat generated by interaction with optical phonons. So the laser has several drawbacks from the practical point of view at present, but the technical community should nonetheless recognize the enormous first step that it represents.

Its key attraction is the prospect it offers of designing the operating frequency of the laser by engineering the 'bandgap'. Leo Esaki and I introduced this idea in 1970 (ref. 3), when we proposed using man-made superlattices as a means of producing terahertz oscillators or amplifiers. Various attempts to produce infrared and light emission followed⁴⁻⁶. Broadly speaking, these schemes fell into two categories: the first is charge oscillations due to Bragg reflection (the well known Bloch oscillation). The other and perhaps more useful type of scheme involves population inversion resulting from injection of carriers via resonant tunnelling into an excited state above the ground state of a quantum well or superlattice.

Using 'lump' rather than 'bulk' structures is a more recent idea^{7,8}. Whenever amplification is present, any slight inhomogeneity may produce fluctuation of charge carriers, and the applied voltage will then become concentrated in regions of lower conductance — known as domains — which are likely to be detrimental to laser action⁹. Breaking up the superlattice into cells separated by highly doped contacts, as here, prevents the formation of domains. Simply put, if Faist *et al.*¹ had not focused their efforts on a cell consisting of a coupled resonant tunnelling structure, it would have been



a, Conduction-band energy of one of the 25 active regions in the quantum cascade laser¹, showing the cell consisting of three coupled quantum wells, and the compositionally graded regions that act as injector and collector. b, Plot of energy E versus transverse momentum

k_t for states in each of the three wells. The bottoms of the parabolas are denoted by the longitudinal energies E_1 , E_2 and E_3 . E_F is the Fermi energy. Carrier relaxations via optical phonons and radiative processes are denoted by τ_{op} and τ_r . (Adapted from ref. 1.)

more difficult to realize lasing.

As I pointed out at the beginning, this multiple-cell laser with frequency determined by band-structure engineering will not only have an enormous impact on potential practical applications, but will also act as a stimulant to researchers engaged in related schemes. Practically any oscillator is potentially capable of lasing with proper cavity feedback and carrier injection. Even the problem of domain formation in 'bulk' devices should not be crucial, as schemes involving bulk structures¹⁰ with domain oscillations may be forced into phase-lock by positive feedback. The reader should be aware that a whole class of electron-transfer devices such as the Gunn oscillator operate on the principle of domain oscillations¹¹.

Finally, to my mind, the frequency of operation is not limited to the near-infrared. With higher energy barriers, there is no reason why the quantum cascade laser could not operate in the visible. To sum up, the new laser has more flexibility in operating frequency but lower efficiency than the standard *pn* junction laser diode because its 'bandgap',

unlike that of the semiconductor diodes, exists only along the coupled quantum wells and thus allows carrier relaxation via fairly efficient phonon interactions. The observation of coherent light emission from the cascade laser demonstrates an important point: the researching of appropriate materials provided by nature is not the only route to the solid-state lasers increasingly demanded by the technical world. □

Raphael Tsu is in the Department of Electrical Engineering, University of North Carolina at Charlotte, Charlotte, North Carolina 28223, USA.

1. Faist, J. *et al.* *Science* **264**, 553–556 (1994).
2. Capasso, F. *et al.* *Appl. Phys. Lett.* **45**, 1193 (1984).
3. Esaki, L. & Tsu, R. *IBM J. Res. Dev.* **14**, 61–65 (1970).
4. Esaki, L., Howard, W. E. & Tsu, R. *US Patent No.* 3,626,328 (1971).
5. Esaki, L. & Tsu, R. *US Patent No.* 4,163,238 (1977).
6. Kazarinov, R. F. & Suris, R. A. *Sov. Phys. Semicond.* **5**, 707–709 (1971).
7. Shmuel, I. *et al.* *Appl. Phys. Lett.* **55**, 654–656 (1989).
8. Kastalsky, A. *et al.* *Appl. Phys. Lett.* **59**, 2636 (1991).
9. Capasso, F., Mohammad, K. & Cho, A. Y. *IEEE Quantum Electron.* **22**, 1853–1869 (1986).
10. Waschke, C. *et al.* *Phys. Rev. Lett.* **70**, 3319–3322 (1993).
11. Shur, M. *Physics of Semiconductor Devices*, Ch. 6 (Prentice Hall, New Jersey, 1990).

IMMUNOLOGY

NF-AT comes under control

Alain Israel

A CRITICAL determinant of the immune response is the ability of the T-lymphocyte antigen receptor rapidly to induce the production of cytokines. In turn, these control events necessary for an effective response¹. The NF-AT transcription complex has been implicated as a nuclear target for the antigen receptor; it is responsible for the synthesis of some of the early activation cytokines, including interleukin 2 (IL-2)², and has therefore been the subject of intense interest.

Last year we had the description³ of the sequence of a partial clone encoding the cytoplasmic subunit of murine NF-AT. Now, with the appearance of the paper by Northrop *et al.* on page 497 of this issue⁴, we have a report of the cloning of the cytoplasmic subunit of human NF-AT. It brings up some unexpected twists: there is, it turns out, more than one type of cytoplasmic subunit, and these subunits probably represent distant members of the well-known Rel/NF- κ B family of transcription factors.

The NF-AT complex is induced fully by the treatment of T cells with the combination of a phorbol ester (such as PMA), which mimics diacylglycerol in activating protein kinase C, and a calcium ionophore (such as ionomycin), which increases levels of intracellular free calcium. NF-AT can be reconstituted from a ubiqui-

tous nuclear component that requires protein synthesis for induction and a T-cell-specific, constitutive cytoplasmic component². This cytoplasmic component (NF-AT_c) translocates to the nucleus in response to calcium signals. The process is inhibited by the immunosuppressants cyclosporin A and FK506, which associate with cytoplasmic receptors (cyclophilin and FKBP, respectively). It is these complexes that constitute the active form of the drugs, and they act by inhibiting the activity of the calcium/calmodulin-dependent protein phosphatase calcineurin⁵. So it seems that calcineurin is somehow involved in the nuclear translocation of the cytoplasmic subunit of NF-AT. This is in accordance with experiments *in vitro* demonstrating that purified NF-AT_c is indeed a phosphoprotein and can serve as a substrate for calcineurin⁶.

In addition, a large body of evidence suggests that the nuclear subunit of NF-AT (NF-AT_n) contains products of the *jun/fos* family, which normally make up the AP-1 transcription factor^{7,8}. This factor is induced in various cell types by stimuli such as PMA, through a mechanism that requires protein synthesis and that is resistant to cyclosporin A. In all, then, full activation of NF-AT results from a PMA-dependent mechanism involving synthesis of the components

of the AP-1 transcription factor, and a calcium-dependent mechanism involving nuclear translocation of NF-AT_c.

The partial complementary DNA encoding the cytoplasmic subunit of murine NF-AT, termed NF-AT_p, was reported by McCaffrey *et al.*³. This protein is expressed specifically in T cells, and can bind to the NF-AT site located in the promoter of the murine IL-2 gene in combination with *c-jun*, or with *c-fos* plus *c-jun*. Moreover, it can activate *in vitro* transcription from a multimerized NF-AT site, when added together with the product of *c-jun*, or that of *c-fos* plus *c-jun*.

Surprisingly, the cytoplasmic subunit of human NF-AT, now cloned by Northrop *et al.*⁴, is not homologous to murine NF-AT_p. However, the sequence of some peptides present in the purified NF-AT_c preparation indicates that it is composed of a mixture of NF-AT_c and NF-AT_p. These data bring into question some widely accepted facts concerning the regulation of NF-AT activity. For instance, unexpectedly, overexpression of NF-AT_c is sufficient to activate transcription of an NF-AT-dependent promoter (for example that of the IL-2 gene) in non-T cells in the presence of the PMA signal, but independently of the calcium signal. And both calcium ionophore and phorbol esters induce changes in the mobility of transfected NF-AT_c: PMA induces an increase in the apparent molecular weight, whereas ionomycin has the opposite effect, suggesting that the two pathways involved in NF-AT activation can both influence its cytoplasmic subunit.

The new work also raises some intriguing questions. First, there are apparently two different cytoplasmic subunits of NF-AT (and maybe more). Are they both present in the same complex, or do several types of complex coexist in the same cell, or do different subsets of T cells contain different types of complex? And are the subunits functionally equivalent, or do they perform different tasks?

Another issue centres on the homology observed between NF-AT_c, NF-AT_p and Rel/NF- κ B family members: if indeed the cytoplasmic subunits belong to the Rel family, even if distantly, then might their cytoplasmic localization be controlled through association with an inhibitory molecule similar to the I κ B- α protein described in the case of NF- κ B (ref. 9)? This molecule prevents access to the nuclear localization sequence of the active NF- κ B subunits; it is degraded in response to different external stimuli, following a modification of its phosphorylation state therefore allowing nuclear translocation of NF- κ B. If there are parallels with I κ B- α , the actual target of calcineurin could be the inhibitor and not the NF-AT cytoplasmic subunit (although the two possibilities are not necessarily exclusive). This could also explain why overexpress-

ion of NF-AT_c bypasses the requirement for ionomycin: titration of a potential cytoplasmic anchoring molecule would result in nuclear translocation in the absence of an external signal.

Finally, the identity of the nuclear subunit NF-AT_n remains unclear. There are strong indications that it contains products of the *jun/fos* family, though the actual identification of these components awaits peptide sequence derived from purified nuclear NF-AT. In this respect, experiments carried out in knockout mice whose T cells are devoid of *c-jun* indicate that they have functional NF-AT activity¹⁰. The physical association with proteins belonging to the *jun/fos* family is not unexpected for members of the Rel/NF- κ B family, as it has been shown that the basic-domain-leucine-zipper region of the *c-jun* and *c-fos* products can directly interact with the Rel homology domain of the p65 subunit of NF- κ B (ref. 11).

We are gradually getting closer to understanding the control of NF-AT activity. That in turn will not only bring insight into some critical aspects of the immune response, but will tell us a great deal about the mechanism of the widely used immunosuppressant cyclosporin A.

Alain Israel is in the Department of Molecular Biology, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France.

1. Crabtree, G. R. *Science* **243**, 355–361 (1989).
2. Flanagan, W. F., Corthesy, B., Bram, R. J. & Crabtree, G. R. *Nature* **352**, 803–807 (1991).
3. McCaffrey, P. G. *et al.* *Science* **262**, 750–754 (1993).
4. Northrop, J. P. *et al.* *Nature* **369**, 497–502 (1994).
5. Liu, J. *et al.* *Cell* **66**, 807–815 (1991).
6. Jain, J. *et al.* *Nature* **365**, 352–355 (1993).
7. Northrop, J. P., Ulman, K. S. & Crabtree, G. R. *J. Biol. Chem.* **268**, 2917–2923 (1993).
8. Jain, J., McCaffrey, P. G., Valge-Archer, V. E. & Rao, A. *Nature* **356**, 801–804 (1992).
9. Beg, A. A. & Baldwin, A. S. *Genes Dev.* **7**, 2064–2070 (1993).
10. Chen, J., Stewart, V., Spyrou, G., Hilberg, F., Wagner, E. & Alt, F. W. *Immunity* **1**, 65–72 (1994).
11. Stein, B. *et al.* *EMBO J.* **12**, 3879–3891 (1993).

NEUROBIOLOGY

A new job for the thalamus

Wolf Singer

ALL routes to the cerebral cortex pass through the thalamus, a collection of nuclei consisting of nerve cells which relay incoming signals to cortical target areas. Each nucleus receives a massive back-projection from the cortical region to which it connects. The function of this feedback loop is thought to be selection of those signals that are most appropriate for subsequent cortical processing, but experimental support for the idea has been hard to come by.

On page 479 of this issue¹ Sillito and colleagues provide evidence that the feedback connections from the visual cortex to the lateral geniculate nucleus (LGN), the thalamic relay for visual signals, synchronize the responses of cells projecting to the visual cortex. Action potentials evoked in these cells by retinal input occur simultaneously much more often than one would expect if the responses of the cells were independent. Furthermore, synchronization is abolished by removal of the visual cortex, which suggests that it is generated by the thalamic feedback projection.

Of particular importance is that synchronization depends on the configuration of the stimuli involved. Synchrony is established only between cells activated by the outline of the same visual stimulus (in this case a straight, moving light bar). If the same pair of cells is coactivated by two spatially segregated, simultaneous light spots, their responses do not synchronize. Cortical neurons respond preferentially to elongated contours of a particular orientation because they re-

ceive converging input from subsets of LGN cells whose receptive fields are aligned along straight lines². Thus synchronization seems to be established selectively when an array of LGN cells is activated by a contour whose orientation matches the preference of those cortical cells which receive convergent input from exactly that array. These data support the hypothesis that cortico-thalamic feedback selects input signals according to the specific requirements of cortical processing^{3,4}. Moreover — and this aspect is particularly fascinating — they shed new light on the idea that selection of neuronal responses and their 'binding' together can be achieved by synchronization⁵.

Many brain functions require neuronal responses to be selected and recombined for further processing. For example, visual scenes evoke patterns of responses on the retina, interpretation of which necessitates a series of highly selective grouping operations. To identify individual figures, neuronal responses evoked by the same object need to be selected and associated with one another, and they must not be confounded with those generated by nearby objects or the background. Related selection and binding functions have to be realized at higher levels of processing, when associations are established between different sensory modalities or sensory representations and motor responses. Because of the large number of possible combinations that these binding mechanisms have to cope with, it is difficult to conceive of hardware solutions which provide a set of converging connec-

tions for every possible constellation. Accordingly, dynamic binding mechanisms have been postulated which allow for a flexible selection and regrouping of responses within fixed neuronal architectures (see ref. 5 for review).

One possible strategy to select subsets of neuronal responses for further joint processing is to raise the discharge rate of the selected neurons. Another way is to make discharge synchronous with that of other cells converging on the same target. These two options differ in important aspects. Increasing discharge rate allows selection of individual responses, but it precludes the possibility of exploiting amplitude modulation for encoding stimulus qualities. Synchronization, on the other hand, always defines relations among responses, as well as making them more salient, and so is equivalent to selecting a particular response constellation. For this second reason, synchronization of neuronal activity could be exploited for the dynamic selection and binding of neuronal responses^{5,6}, and evidence from multielectrode recordings in the cerebral cortex supports this possibility⁵.

Sillito *et al.* now show that response selection and binding by synchronization may also apply in the thalamus. But what type of selection and binding problems might arise at this low level of processing? To generate orientation-selective cortical neurons that cover the whole range of distinguishable orientations, the output of a given group of LGN cells needs to be distributed in ever-changing combinations to a large number of cortical cells. This can be achieved by a flexible binding mechanism. By synchronizing only the responses of those LGN cells that are aligned along a particular contour, cortical cells can flexibly select the combination of input signals that best matches the stimulus constellation. If the latter changes, subsets of the same cells can be selected and bound with others within unchanged hardware. So highly selective input configurations can be defined in functional terms without requiring matching selectivity of anatomical connections, and this could much reduce the required number and precision of thalamo-cortical connections.

Related binding problems are inherent in coarse coding. Information about the exact position and orientation of a contour line is represented by the graded responses of a large population of broadly tuned cortical cells⁷. This economizes on neuron numbers, because different stimuli can be represented by varying relations among a limited set of active neurons, and one need not implement an individual sharply tuned cell for every distinguishable position and orientation. But population codes pose binding problems if several, nearby contours need to be represented at the same time. Out of the many cortical responses, those subsets which

represent the same contour have to be selected for joint processing. Again, if LGN cells responding to the same contour can be distinguished by virtue of synchronous firing, this signature can be exploited at the cortical level for the flexible association of population responses coding for the same contour.

An intriguing question is how cortical feedback can override the timing of discharges imposed by retinal input. Comparison with cortical synchronization phenomena may provide an answer. These also occur with zero phase lag and are transitory, independent of external timing and strongly influenced by stimulus configuration. Synchronization of cortical responses is achieved by reciprocal cortico-cortical connections⁸. These closely resemble the cortico-thalamic projection, in that they are glutaminergic and contact both excitatory and inhibitory neurons within the respective target structure. Simulation studies of cortical synchronization suggest that the emergence of coherent firing can be understood as the result of a dynamic, self-organizing process based on re-entry (see ref. 9 for review).

Thalamic synchronization might therefore be considered as an integral part of selection and binding operations that occur simultaneously at various levels of processing, and that influence one another through re-entry loops. The finding that thalamic synchronization emerges only during later phases of the light response agrees with this view. Together with the fact that cortico-thalamic fibres considerably outnumber thalamo-cortical projections, it constitutes strong evidence against the classical notion that the thalamus simply serves as a gateway through which signals must pass to the cortex (one which is shut during sleep, and opened during wakefulness). The organization of the various thalamic nuclei is strikingly similar, and all cortical areas have feedback connections with thalamic nuclei. So synchronization may well be exploited as a mechanism for dynamic selection and binding of responses throughout the forebrain, and may not be restricted to sensory processing alone. □

Wolf Singer is in the Max Planck Institute for Brain Research, Deutschordenstrasse 46, D-60528 Frankfurt/Main, Germany.

Actors, spectators and control

David J. Tanner

A CONCEPTUALLY simple approach to controlling the course of chemical reactions is to excite, in one of the reactants, a vibration that is directed along the reaction coordinate of the desired bimolecular reaction. Indeed, in the past few years there have been several fascinating experiments¹⁻³ on the reaction of H atoms

in the CN product, suggesting that the C-N bond is not a spectator but rather a participant in the reaction.

There is additional evidence for the participation of the C-N bond in the reaction of Cl + HCN. The hot vibrational energy distribution in the CN product is virtually independent of the vibrational

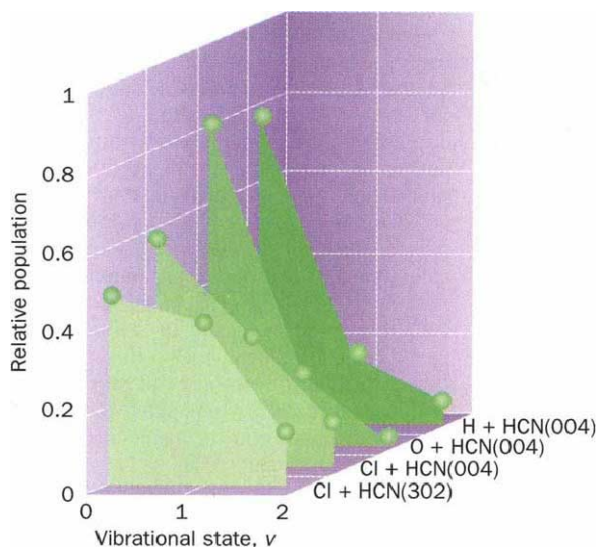


FIG. 1 Vibrational state populations of the CN product, for different reactions $X + \text{HCN}$. Note that the product distribution is almost the same for Cl + HCN whether the HCN originates from the (004) or (302) vibrational state. (All figures courtesy of F. F. Crim and J. Pfeiffer.)

with vibrationally excited HOD. If the O-H bond is excited there is a marked tendency to form $\text{H}_2 + \text{OD}$; if the O-D bond is excited the propensity is for $\text{HD} + \text{OH}$. Water, with its isolated high-frequency vibrations, is almost an ideal candidate for this type of bond-selective chemistry: the nonreactive bond is a spectator to the reaction, receiving none of the energy released while retaining all that it initially possesses.

At a conference this spring*, Metz, Pfeiffer, Thoenke and Crim (University of Wisconsin) reported that they had extended this approach to another molecule, HCN, whose C-H stretching vibration should become the reaction coordinate in hydrogen abstraction reactions, $X + \text{HCN} \rightarrow \text{HX} + \text{CN}$ (see also ref. 4). When the attacking atom, X, is H or O the CN product is vibrationally cold, indicating that it is indeed a spectator in the abstraction reactions with these atoms. When the attacking atom is Cl, however, the situation is very different (Fig. 1). Reactions with chlorine atoms produce considerable vibrational excita-

tion in the CN product, suggesting that the C-N bond is not a spectator but rather a participant in the reaction. Whether the initial vibrational excitation is deposited as four quanta of C-H stretch (004 state), or two quanta of C-H stretch and three of C-N stretch (302 state), the vibrational energy distribution of the CN fragment is the same, within 10 per cent (Fig. 1). In the first case the C-N bond receives energy from the C-H bond; in the second case it donates energy. The fact that the outcome is virtually independent of initial conditions is the hallmark of a statistical or chaotic process. This is evidence for the formation of a HCICN complex in addition to, or instead of, the direct abstraction of hydrogen.

A clue to understanding the difference between the reactions Cl + HCN and H or O + HCN comes from appreciating the variety of reactions possible for the latter. The reactions of H or O atoms with vibrationally excited HCN can form products such as NCO, probably by initial attack on the multiple bond and subsequent decay of a transient complex (Fig. 2). Thus, the direct abstraction reaction produces one product and the indirect reaction through a complex produces another. The energetics of the Cl com-

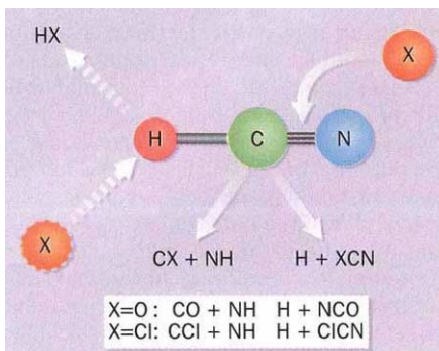


FIG. 2 Schematic diagram showing the attack of X on HCN for the direct hydrogen abstraction reaction and for complex formation at the CN site.

- Sillito, A. M., Jones, H. E., Gerstein, G. L. & West, D. C. *Nature* **369**, 479-482 (1994).
- Chapman, B., Zahs, K. R. & Stryker, M. P. *J. Neurosci.* **11**, 1347-1358 (1991).
- Crick, F. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4586-4590 (1984).
- Sherman, S. M. & Koch, C. *Expl Brain Res.* **63**, 1-20 (1986).
- Singer, W. *A. Rev. Physiol.* **55**, 349-374 (1993).
- von der Malsburg, C. & Schneider, W. *Biol. Cybern.* **54**, 29-40 (1986).
- Lehky, S. R. & Sejnowski, T. J. *J. Neurosci.* **10**, 2281-2299 (1990).
- Engel, A. K., König, P., Kreiter, A. K. & Singer, W. *Science* **252**, 1177-1179 (1991).
- Tononi, G., Sporns, O. & Edelman, G. M. *Cerebral Cortex* **2**, 310-335 (1992).

* American Chemical Society Meeting, San Diego, California, USA, 13-18 March 1994.

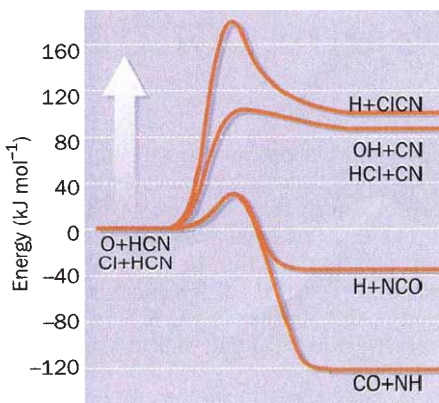


FIG. 3 Schematic potential energy profiles for the reactions $O + HCN$ and $Cl + HCN$. The high energy of the $ClCN$ product (and even higher energies of CCl and NCI , not shown) limit the possible outcomes of $H-CN$ complex formation.

plex, however, seem to prohibit the alternative reaction: the barrier to forming $H + ClCN$ would appear to be larger than the total available energy (Fig. 3). Thus for Cl atom attack at the CN the complex can only decompose back to reactants or form $HCl + CN$ — the same products formed in the direct reaction. The formation of this complex means that the $C-N$ bond is no longer a spectator but rather a full participant, receiving its share of the available energy in the reaction.

These measurements demonstrate both the simplicity and the subtlety of vibrational state control of chemical reactions. Even exciting a well isolated vibration does not guarantee successful control: although the $H-CN$ bond is still cleaved in this reaction, one sees from these results that, in general, attack at other bonds will compete with the desired reaction path. One needs the right excitation and the right chemistry for vibrational state control to be successful.

The limitations of vibrational state control should in no way be viewed as discouraging. These detailed measurements are providing portions of the complete map of state-to-state quantum transition amplitudes: starting in a particular initial quantum state, what is the probability of ending up in a particular final quantum state? Detailed theoretical simulations of these reactions can provide insight into the intimate details of the dynamics of the complex, and refine our understanding of the potential energy surfaces on which the reaction takes place⁵⁻⁷. Given complete knowledge of the potential energy surface and the state-to-state dynamics, one can contemplate augmenting vibrational state control with other techniques to achieve selectivity of chemical products. A macroscopic analogy might be the control of an automobile (where inelastic collisions and undesired fragments are to be avoided also). Vibrationally unselected reactions follow no planned route. Vibrational state

control sets a well defined route which, once initiated, cannot be changed. The control of the future can be compared to automated driving, in which a computer maps many routes and instantaneously adjusts the vehicle direction and speed depending on traffic conditions (and police surveillance).

One good way forward might be to combine vibrational state control with oriented molecules⁸⁻¹¹. The relative orientation of reactants strongly affects which products are formed; indeed, in all the $X + HCN$ reactions, it is probably the angle of attack that determines whether direct abstraction or complex formation occurs. Relative orientation is probably the single most important feature not controlled in the vibrational-state-controlled experiments. In addition, one might aim to prepare more complicated vibrational excitations, including coherent 'superposition' states, which lead selectively to the desired chemical product. Such states have been proposed for selective bond breaking in unimolecular reactions¹²⁻¹⁵. Technically, this is much more difficult for bimolecular reactions: it would essentially involve synchronization of vibrational preparation and collision on a femtosecond timescale.

Despite the practical difficulties that lie ahead, there is widespread optimism that control of bimolecular reactions will become a reality. In the same way that microwaves are replacing (or at least supplementing) conventional ovens, it is not difficult to imagine that fifty years from now the present paradigms for doing chemistry — mixing substances in solution together and heating — will seem primitive and quaint. □

David J. Tannor is in the Department of Chemistry and Biochemistry, University of Notre Dame, Indiana 46556, USA.

Virtual assembly

THE current elections for the European Parliament fill Daedalus with gloom. Parliaments, he points out, no longer decide anything. These days they are simply devices for judging the emotional ascendancy of the competing parties, like the screaming matches conducted between rival tribes of monkeys. The European Parliament cannot even carry out this function. Multiple translation obscures the mood of the assembly. It just reflects the deeper absurdity of trying to unite nations that speak many different languages.

Modern technology, says Daedalus, can do much better. Years ago he proposed a 'virtual committee' whose members did not meet, but typed their remarks into a common computer system which distributed them to all the screens. Cunningly, all contributions were anonymous. Leaders could not enforce their opinions, and underlings could not simply support their own side. All were forced to consider the proposals on their merits. The changing mood of the virtual committee was monitored by frequent polling and voting. Systems like this are now in experimental operation. Any number of people can 'talk' at once, so shy geniuses cannot be shouted down by forceful blusterers. All comments stay on the record, visible to all parties; minutes are kept automatically.

Clearly this is the way forward for the European Parliament. The expensive buildings in Strasbourg and Brussels should be sold, the endless travelling eliminated, and the whole thing conducted on the Internet. Machine translation could provide rough drafts of each comment in every official language; human translators would tidy these up and release them onto the Net.

A totally new political process should result. The traditional tribal warfare, the dishonest rhetoric and tedious oratory, would be replaced by thoughtful written arguments. Unsocial parliamentary hours would cease. Debates would run continuously and in parallel. Each member could log off at any time, and return later to scan the accumulated discussion, add his own comments and vote on current resolutions. He could consult expert witnesses elsewhere on the Net. The flood of facts and arguments might even make him change his mind! The 'Delphi' technique of repeated voting, with all voters seeing the results instantly, leads to rapid consensus: sound, well argued policies should emerge at a great pace. To foil impersonators, and to free the members from having to play to a gallery of electronic eavesdroppers, the debate should be encrypted.

David Jones

1. Sinha, A., Hsiao, M. C. & Crim, F. F. *J. chem. Phys.* **92**, 6333–6335 (1990); **94**, 4928–4935 (1991).
2. Metz, R. B., Thoemke, J. D., Pfeiffer, J. M. & Crim, F. F. *J. chem. Phys.* **99**, 1744–1751 (1993).
3. Bronikowski, M. J., Simpson, W. R. & Zare, R. N. *J. phys. Chem.* **97**, 2194–2203 (1993); **97**, 2204–2206 (1993).
4. Metz, R. B., Pfeiffer, J. M., Thoemke, J. D. & Crim, F. F. *Chem. Phys. Lett.* **221**, 347–352 (1994).
5. Bair, R. A. & Dunning, T. H. Jr. *J. chem. Phys.* **82**, 2280–2294 (1985).
6. Brooks, A. N. & Clary, D. C. *J. chem. Phys.* **92**, 4178–4190 (1990).
7. Sun, Q. & Bowman, J. M. *J. chem. Phys.* **92**, 5201–5210 (1990).
8. Gandhi, S. R. & Bernstein, R. B. *J. chem. Phys.* **87**, 6457–6467 (1987).
9. Greeley, J. N., Martin, J. S., Morris, J. R. & Jacobs, D. C. *Surface Sci.* (in the press).
10. Judson, R., Lehmann, K. K., Warren, W. S. & Rabitz, H. A. *J. molec. Structure* **223**, 425–456 (1990).
11. Friedrich, B. & Herschbach, D. R. *Nature* **353**, 412–414 (1991).
12. Brumer, P. & Shapiro, M. A. *Rev. Phys. Chem.* **43**, 257–282 (1992).
13. Kosloff, R., Rice, S. A., Gaspard, P., Tersigni, S. H. & Tannor, D. J. *Chem. Phys.* **139**, 201–220 (1989).
14. Warren, W. S., Rabitz, H. & Dahleh, M. *Science* **259**, 1581–1589 (1993).
15. Krause, J. L., Whitnell, R. M., Wilson, K. R., Yan, Y. J. & Mukamei, S. J. *chem. Phys.* **99**, 6562–6578 (1993).

Single sweetness signal

SIR — The number of neural signals encoding information (coding dimensionality¹) about sweet-tasting compounds is vigorously debated^{2,3}. One hypothesis is that many signals are involved at every level of coding; accordingly, each physically distinct, sweet-tasting compound binds to a distinct population of receptors, producing a unique pattern of activity at the receptor level and all subsequent levels of coding⁴. An alternative hypothesis is that only one or a few signals encode information about sweet-tasting compounds⁵⁻⁷.

To investigate the coding dimensionality of three simple saccharides, α -D-glucose, α -D-fructose and sucrose (α -D-glucopyranosyl- β -D-fructofuranoside), we used a two-alternative, forced-choice taste discrimination protocol. In each ses-

sion a subject was required to discriminate a standard stimulus, S , from a test stimulus, T , that varied in concentration (all solutions were allowed to auto-rotate for 24 hours). The figure shows the data from six subjects in which $S = 200$ mM fructose or sucrose and $T = [\text{glucose}]$, varied over the concentration range 200 mM to 1.1 M. In each case the probability of discrimination, $P_{\text{discriminate}}(T/S)$, was found to be a U-shaped function of T , declining approximately to chance. Similar U-shaped discrimination functions were found in six other similar experiments (data not shown) in which $S = 100$ mM fructose or $S = 100$ mM sucrose and $T = [\text{glucose}]$.

To determine if discriminability declines to chance, we used a χ^2 statistic to test the hypothesis $P_{\text{discriminate}}(T/S) = 0.5$ for each pair (T, S). All the experiments described in the figure yielded at least one point that does not reject the hypothesis ($P \geq 0.1$). For the complete set of 12 experiments with 100 or 200 mM fructose or sucrose standards, 10 had points near or at the minima that failed to reject ($P \geq 0.1$); the remaining two experiments had one or more points that failed to reject ($0.05 < P < 0.1$). Because the data points that failed to reject the hypothesis were always near the minima of the discrimination functions, and because there were always points on both sides of the minimum that very strongly rejected the hypothesis, the failures to reject do not result from inadequately sensitive experimental methodology or from an overall lack of statistical power, but rather must arise from a loss of discriminability.

To characterize quantitatively the discrimination functions, we fit each set of data with a two-limbed 'inverted gaussian' function described by the equation. T is the test stimulus concentration, μ is the concentration at which discriminability of test and standard is a minimum, δ is the depth of the curve and characterizes the loss of

discrimination, and $\sigma^\pm$ is the steepness of the sides of the function. We conclude that, within the concentration ranges tested, glucose, fructose and sucrose are indeed indistinguishable when their relative concentrations are suitably adjusted, and therefore that normal subjects are

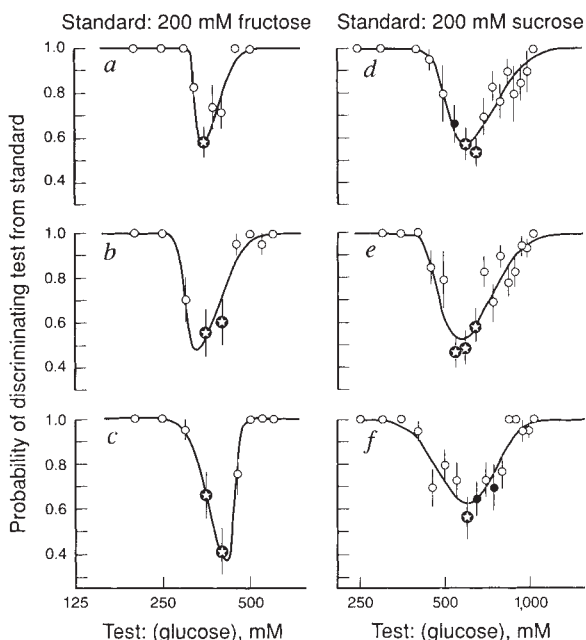
$$P_{\text{discriminate}}(T/S) = 1 - \delta \exp \left[-\frac{1}{2} \left(\frac{\log T - \log \mu}{\log \sigma^-} \right)^2 \right], \quad T \leq \mu$$

$$= 1 - \delta \exp \left[-\frac{1}{2} \left(\frac{\log T - \log \mu}{\log \sigma^+} \right)^2 \right], \quad T > \mu$$

monogamous for these saccharides.

In light of this conclusion, the ascending and descending limbs of the discrimination functions in the figure can be viewed as psychometric functions for increment and decrement thresholds. With the 70% level as the threshold criterion, the Weber fractions obtained from the equation satisfy $\pm \Delta T/T = 2.3 \log \sigma^\pm$, where $T = \mu$ and $\pm \Delta T$ is the increment/decrement threshold. Pooling all 24 estimates of $\log \sigma^+$ and $\log \sigma^-$ ($\delta = 0.5$) across experiments 1-12, we obtain $|\Delta T|/T = 0.16 \pm 0.01$, that is, a 16% Weber fraction. These values compare well with the average reported values ($\sim 20\%$) for sucrose⁸.

The finding of monogamia implies that under the present experimental conditions the three sugars investigated act identically somewhere along a common sensory pathway⁹. Similar arguments have emerged from the study of photopic (day vision) and scotopic (night vision) light matching. In observers with normal colour vision, for example, a light of any spectral distribution can be made indistinguishable by a suitable mixture of three 'primaries'¹. Thus, day vision is trichromatic, and the coding dimensionality of photopic stimuli is three. Under scotopic conditions, any two lights can be rendered indistinguishable by an adjustment of their relative intensities. Thus, in night vision, observers are monochromatic, and the coding dimensionality for spectral information is one¹⁰. Scotopic monochromacy arises because only rod photoreceptors produce signals that reach



Each panel is a discrimination function for one subject for pairs of simple sugars. *a-c*, fructose; *d-f*, glucose. In the two-alternative forced-choice procedure each well trained subject was given 10 discrimination trials per day. For each function and across sessions a single taste standard, S , was used, while the test stimulus, T , was varied. Each trial consisted of successive whole-mouth tasting of one of four triplets of stimuli: $\langle SST \rangle$, $\langle TSS \rangle$, $\langle STT \rangle$, $\langle TTS \rangle$. The observer's task was to choose whether the first or third stimulus was the odd stimulus. Each data point (~ 50 trials per minima) represents $P_{\text{discriminate}}(T/S)$, the estimated probability for the subject to discriminate T (glucose), given by the abscissa, from S , identified above the panel. The discrimination probability is the fraction of correct trials; the error bars are $\pm$ s.e.m. Starred circles are not significantly different from chance, having significance levels $P \geq 0.1$ (χ^2 test); Filled circles are weakly different from chance, with $0.05 \leq P < 0.1$; open circles are strongly different from chance, with $P < 0.05$. The hypothesis that $P_{\text{discriminate}}(T/S) = 0.5$ was tested for each data point with a χ^2 test, assuming the data are Bernoulli trials. Parametric results of a curve fitting/likelihood-ratio statistic support these conclusions (E. N. Pugh, Jr *et al.*, unpublished data).

1. Wyszecki, G. & Stiles, W. S. *Color Science* (Wiley, New York, 1982).
2. Bartoshuk, L. M. In *Sweetness* (ed. Dobbing, J.) 33-43 (Springer, New York, 1987).
3. Beidler, L. M. & Tonosaki, K. In *Taste, Olfaction and the Central Nervous System* (ed. Pfaff, D. W.) 47-64 (Rockefeller Univ. Press, New York, 1985).
4. Erickson, R. P. *Physiol. Behav.* **28**, 57-62 (1982).
5. von Sramlik, E. *Handbuch der Physiologie der Niederen Sinne Die Physiologie des Geruchs- und Geschmackssinnes I* (1926).
6. Shallenberger, R. S. *Taste Chemistry* (Blackie, London, 1993).
7. Birch, G. G. In *Sweetness* (ed. Dobbing, J.) 3-14 (Springer, New York, 1987).
8. Pfaffmann, C., Bartoshuk, L. M. & McBurney, D. H. In *Handbook of Sensory Physiology v. IV, Chemical Senses Part 2: Taste* (ed. Beidler, L. M.) 75-101 (Springer, New York, 1971).
9. Teller, D. Y. & Pugh, E. N. Jr in *Colour Vision* (ed. Mollon, J. D.) 577-589 (Academic, London, 1983).
10. Pugh, E. N. Jr in *Steven's Handbook of Experimental Psychology Vol. 1: Perception and Motivation* (eds Atkinson, R. C. *et al.*) 75-164 (Wiley, New York, 1988).

higher visual centres in night vision; thus any two light stimuli causing equal rates of rhodopsin isomerization produce indistinguishable neural signals¹⁰. An explanation for the monogeusia analogous to that accounting for scotopic monochromacy is that the three sugars bind reversibly to a single class of membrane receptor in taste cells. On the assumptions that such binding is first-order, and that the psychophysical indiscriminability relation is transitive, our data imply that at the tongue

surface the apparent relative affinities for sucrose, fructose and glucose are 1, 2/3 and 1/3, respectively.

P. A. S. Breslin

S. Kemp*

G. K. Beauchamp

*Monell Chemical Senses Center,
Philadelphia,
Pennsylvania 19104,
USA*

*Present address: Givaudan-Roure, Teaneck, New Jersey 07666, USA.

Climate effects on mountain plants

SIR — Temperature-limited environments such as boreal regions, arctic regions and high mountains are thought to be very sensitive to greenhouse warming¹. As a result of such warming, for example, plant species and communities in high mountains will be pushed upwards in elevation and may be eliminated if already at mountain summits^{2,3}. If evidence were to be found of a continuing change of summit floras, this would indicate that global warming is already having a significant ecological impact. Although preliminary observations have recently been presented⁴, no conclusive evidence has

yet been reported.

During the summer of 1992, we collected data on the state of the flora at 26 summits exceeding 3,000 m in the middle part of the Alps (western Austria, eastern Switzerland) and compared the actual records on cover and abundance of vascular plant species with historical records^{5–10}. Species richness has increased during the past few decades (see figure), and is more pronounced at lower altitudes. However, upward movement of the alpine-nival flora is an overall trend, as indicated by the comparison of the fitted regression lines for species-richness with

altitude. Although the presented data are not corrected for age of the historical record or for the expansion of the considered belt, the difference between the two regression lines is obvious. The exponential decrease of species richness with altitude is a general feature of the nival vegetation above the closed alpine grasslands.

On the basis of 12 very precise historical records^{8–10}, we calculated moving rates for nine typical nival plant species. The maximum values approached 4 m per decade, although most of the values were below 1 m per decade. The old records used for this calculation were collected about 70–90 years ago. According to the records from meteorological stations in Austria, the mean annual temperature has increased since that time by 0.7 °C (ref. 11). Taking into account an average decrease of 0.5 °C per 100 m of increasing altitude, this warming should theoretically lead to a shift in the altitudinal vegetation belts at a rate of 8–10 m per decade. The empirical

values are clearly far lower than this.

In conclusion, there is no doubt that even moderate warming induces migration processes, and that this process is under way. The example from the limits of plant life at high alpine summits is of general importance and suggests that global warming is already having a significant effect on alpine plant ecology. Even in situations where plants must move upwards, the warming is sufficient to stimulate migration, and may cause disastrous extinctions in these environments.

Georg Grabherr, Michael Gottfried

Harald Paul

Department of Vegetation

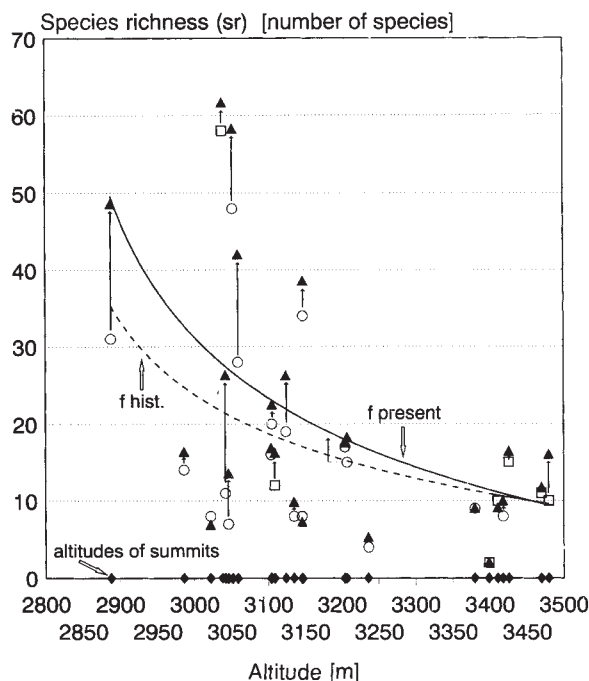
*Ecology and Conservation Biology,
University of Vienna, PO Box 285,
A1091 Vienna, Austria*

1. Markham, S., Dudley, N. & Stoltton, S. *Some Like it Hot* (WWF International, Gland, 1993).
2. Holten, J. I. in *Impacts of Climate Change on Natural Ecosystems* (eds Holten, J. I. et al.) 84–105 (Norwegian Institute for Nature Research, Trondheim, 1993).
3. Nilson, S. & Pitt, D. *Mountain World in Danger* (Earthscan, London, 1991).
4. Hofer, H. R. Ber. Geobot. Inst. ETH, Stiftung Rübel **58**, 39–54 (1992).
5. Klebelsberg, R. *Das Vordringen der Hochgebirgsvegetation in den Tiroler Alpen*. *Österr. Bot. Zeitschr. Jahrg.* 177–186; 241–254 (1913).
6. Rübel, E. *Pflanzengeographische Monographie des Berninagebietes*. *Bot. Jahrbuch Bd.* 47/1–4 (1911).
7. Reising, H. & Pitschmann, H. *Obere Grenzen von Flora und Vegetation in der Nivalstufe der Zentralen Ötztaler Alpen (Tirol)*. *Vegetatio*, **VIII**, 93–129 (1958).
8. Braun-Blanquet, J. *Ein Jahrhundert Florenwandel am Piz Linard (3414 m)* (Bull. Jard. Bot., Bruxelles, 1957).
9. Braun, J. *Die Vegetationsverhältnisse in der Schneestufe der Rätisch-Lepontischen Alpen* **48**, 156–307 (Neue Denkschr. Schweiz. Naturforsch. Ges., 1913).
10. Braun-Blanquet, J. *Über die obersten Grenzen Pflanzlichen Lebens im Gipfelbereich des Schweizerischen Nationalparks* **6**, 119–142 (Kommission der Schweizerischen Naturforsch. Ges. zur Wiss. Erforsch. des Nationalparks, 1958).
11. Auer, R. et al. in *Bestandsaufnahme anthropogene Klimaänderungen mögliche Auswirkungen auf Österreich – mögliche Massnahmen Österreich* (Österr. Akademie Wissenschaft, 1992).

Scaling and brain connectivity

SIR — From published information describing the connections between areas of macaque monkey visual cortex, Young has applied non-metric multidimensional scaling (NMDS)^{1–4} to construct a two-dimensional map of 30 areas⁵. Areas are represented as points and were found to be arranged on a circle, which Young took as evidence for the 'two-streams' hypothesis of visual processing.

NMDS is designed for ordinal data. From a matrix of similarities between a given set of objects, a representation of the objects in a geometric space is sought such that the rank order of the distances between objects in the space corresponds to the reverse rank order of the similarities. In Young's case, similarity encodes the degree of connectivity for each pair of areas, with the connectivity specified as either reciprocal, unidirectional, or absent. We first explored the fit of his



Species richness of historical and our present-day records at nival summits in the Alps plotted against altitude. Rare species are downweighted (0.25), species of moderate abundance were given the weight 0.5, frequent species the weight 1.0. Rare species were downweighted due to the comparatively high probability that these were overlooked by the original authors. Displayed are 24 summits with siliceous bedrock. Age of the historical records is 40 to 90 years (circles, 1895–1918; squares, 1947–53; triangles, present-day). Arrows, increase in species richness, which is pronounced at lower altitudes and low to moderate at higher altitudes. Most summits considered are climbed only occasionally, so the effect of humans is negligible.

configuration to the original data. One goodness-of-fit measure, RSQ, is the proportion of variance in the similarities (after a monotonic transformation) accounted for by the distances. Young quotes an RSQ of 40% (see ref. 5), whereas only values close to 100% are usually regarded as acceptable. It could be that the small number of similarity levels prevents the RSQ from reaching 100%. We found by analysis (to be reported elsewhere) that the maximum value of RSQ attainable is indeed data-dependent. For the map found by Young⁵ using the secondary approach to tied similarities³, the maximum RSQ value is 76% in the case of binary similarity data — ternary data will have an even higher RSQ ceiling. This makes Young's value of 40% more respectable, but still the fit is not good.

We next looked at how low fit could compromise the form of a computed map when using the secondary approach to ties. To generate very low fit configurations, we produced two-dimensional NMDS representations of ternary similarity matrices in which the entries were assigned at random. This yields circular configurations, with low RSQ, a finding consistent with theoretical considerations⁶. (As those RSQ values lie below 0.4, the hypothesis that the visual system data is totally due to a random process can be rejected. However, such arguments do not guarantee that a derived structure is free of artefacts⁷.) We also ran NMDS on dissimilarity matrices generated from a set of distances between points in the plane which had been corrupted by noise to weaken the fit. With zero noise, the structure is recovered and the RSQ is high. As more noise is added to the distances, the RSQ gets lower and the derived structure becomes more circular. Finally we applied NMDS to adjacency data from rectangular grids. Again, the fit is poor and circular configurations result.

If the circularity of Young's configuration genuinely reflects underlying structure, then the connectivity data possesses serial order (with wraparound). We therefore tested explicitly for this by taking a standard non-NMDS method⁸ designed for finding serial order and tailoring it to the wraparound case⁹. If the rows and columns of the connectivity matrix can be arranged so that the non-zero entries cluster on the diagonal and corners of the matrix, then wrapped serial order is present and is specified by the row assignment. Using Young's convention that an uninvestigated connection be treated as absent, we examined the binarized connectivity matrix arranged in the order implied by his map. There is some serial order, though many connections are displaced from the appropriate regions. The degree of departure from perfect ordering can be quantified by regarding each row vector as a point in a high dimensional

space and finding the shortest circuit between the points so generated⁸. Using a standard heuristic the shortest length we found was 206, whereas the circuit implied by Young's configuration has length 240. In contrast, the theoretical minimum is 60 for perfect serial ordering with wrap-around, and an average random circuit has length 437.

Application of NMDS to the visual system, with similarity encoding the nature of the direct connections, produces a poor fit configuration. In cases of poor fit there is a bias towards circular structures, whether or not there is genuine serial order. It is essential to be aware of this in interpreting the structures derived. Our results suggest that Young's configuration reflects a mixture of both genuine and artefactual structure. Most of the other connectivity datasets to which he has applied his method have also yielded circular or near-circular configurations^{10,11}. This striking similarity across systems points to either a deep principle of brain organization, or an artefactual component in the method.

Martin W. Simmen

Geoffrey J. Goodhill

David J. Willshaw

*Centre for Cognitive Science,
University of Edinburgh,
Edinburgh EH8 9LW, UK*

YOUNG *ET AL.* REPLY — Simmen *et al.* suggest that NMDS introduced artefactual structure because the fit of the solution for the primate visual system⁵ was not close to 100%. We have tested this idea quantitatively¹², by examining various arbitrary multidimensional structures. When given the metric distances between the vertices of these structures, NMDS recovers their parameters with very high RSQ. When we degraded the data to the same level of measurement as connectivity data, by setting thresholds for proximity between vertices, NMDS produces solutions with RSQ typically between 0.2 and 0.5. We then tested whether the degraded-data solutions recovered the known structure of the input data. Using Procrustes rotation^{5,13}, we found that the fit between the structures derived from degraded data and the metric structures is excellent, typically yielding correlations around 0.95. Therefore, contrary to the suggestion of Simmen *et al.* above, serious artefact is not introduced by NMDS into solutions with RSQ comparable to that of the reported one, and the RSQ of the reported solution therefore cannot itself be interpreted as evidence for the presence of artefactual structure¹².

The traditional means of deciding whether a solution should not be trusted to be a systematic reflection of data structure is to compare the RSQ of a solution with analysis of comparable random data, thereby determining whether the solution

is one in which low fit could have produced an artefact¹⁴. We examined the distribution of RSQ statistics from scaling 200 randomly re-ordered matrices. The probability of the RSQ of the reported solution falling within this low fit distribution is less than 10^{-30} . We found that the recovery of known structure from low level test data is badly compromised only when solutions fall in the low fit distribution of RSQ for those data. We note that Simmen *et al.* do not indicate either the RSQs or the relation of these statistics to the low-fit distributions for their artefactually circular test data and that their artefactual results are consistent with their solutions falling within the low-fit distributions, unlike the visual system solution.

A circular configuration can arise artefactually in a two-dimensional NMDS solution when fit is very poor. Data structure, however, can itself be circular, as in Shepard's famous analysis of colour judgement data². These two interpretations can be further distinguished by deriving higher dimensional solutions, in which artefactual structure takes a spheroidal or hyperspheroidal form, and systematic structure should still take a planar circular form. Accordingly, we examined solutions in up to six dimensions, which have higher RSQ (0.7 in the six-dimensional case). A near-planar circular structure would be expected if the data structure reflects a system that is divided into two gross streams¹⁵, which possess a common origin in occipital cortex^{16,17} and reconverge after somewhat segregated processing (see, for example, ref. 18). All higher dimensional solutions are statistically significantly related to the orderings of the areas derived from independent analyses by others^{15,16} that embody these organizational principles⁵ (every $p < 0.000001$). The forms of all the higher dimensional solutions were near-planar, and none was spheroidal. Hence, the form of the solution is a systematic reflection of underlying data structure and is not due to artefact. Indeed, these significant correspondences constitute external validation, and would be very difficult to explain if there were any serious perturbation of the NMDS solution by artefact.

Simmen *et al.* use a non-NMDS method, normally used to examine the chronological order of grave sites⁸, to assess whether there is serial order in the data. But this analysis makes the assumption that there is a single dimension, time in the application for which the method is designed, underlying the data. This assumption is breached by these data: time does not possess parallel streams between which there can be cross-talk. Lateral cross-talk connections will be displaced from the appropriate regions and elevate the circuit length. Further, the visual system is a level-spanning

hierarchy¹⁶, in which many structures make connections with much higher or lower areas. These level-spanning connections contribute to the NMDS structure, but form a pattern that is not accommodated by the seriation method, so elevating the circuit length. Despite these difficulties, we optimized circuit length explicitly with simulated annealing and found an ordering with shortest circuit length of 196. There certainly is serial order in the data: the probability that the visual circuit falls in the distribution of random circuits is vanishingly small ($p < 10^{-49}$), and the data are more strongly serial than most empirical datasets in Archeology that are accepted as (serially ordered). Statistical comparison of the arrangement of areas in the optimal length ordering with the NMDS solution yielded a correlation of 0.9 ($p < 0.000001$). Parietal and temporal areas were maximally segregated, being joined at the one side by striate and prestriate areas, and at the other by STP and area 46. Hence, this method further corroborates the NMDS result. Simmen *et al.* do not explain how the excellent correspondence between four independent methods could come about by blind artefact.

The NMDS structures reported for all central sensory systems are curved. This does indeed reflect a principle of brain organization, though not a deep one: central sensory systems are all sparsely connected^{12,15}. It is well-known that NMDS analysis of sparse similarity matrices produces curved structures¹⁹. NMDS faithfully reflects this *bona fide* aspect of data structure as it does more interesting ones.

We conclude that: (1) analysis of test

data shows that solutions with fit in the range of the reported one can recover the structure of known input data almost perfectly; (2) the solution is not a low-fit one in which circular structure is likely to emerge artefactually; (3) independent analyses externally validate the result; (4) this validation extends to higher-dimensional solutions that would not have a planar circular form due to artefact. NMDS remains a potent means of analysing central nervous connectivity, which has already helped to identify the site of novel physiological features²⁰.

Malcolm P. Young

Jack W. Scannell

Gully A. P. C. Burns

Colin Blakemore

*University Laboratory of Physiology,
Parks Road,
Oxford OX1 3PT, UK*

Regional variation in fruitflies

SIR — Begun and Aquadro¹ estimate DNA variation, using four-cutter enzyme restriction analysis of genomic DNA, in two populations of *Drosophila melanogaster* from North Carolina and Texas, and one from Zimbabwe, and have compared their results on genetic structure with our published results. They conclude: (1) that the African population is more variable at the DNA level than those from the United States; (2) that most DNA restriction variants are not shared between the two geographical regions; and (3) that there is an unappreciated degree of population structure in *D. melanogaster* and the equilibrium models of molecular evolution are inappropriate for this species.

We wish to point out that although the African population studied in ref. 1 comes from a different region (East Africa) than those studied in our laboratory (West and Central Africa), we have reached similar conclusions using four-cutter restriction analysis of mitochondrial DNA variation²⁻⁴. The mtDNA variation data from a worldwide collection of lines show: (1) that intra-population diversity is about twice as high as that for the protein variation; (2) that population differentiation, as measured by the fixation index (F_{ST}), is roughly five times higher than that for a protein variation; and (3) that geographical populations of *D. melanogaster* harbour population and/or region-specific mtDNA haplotype(s).

Higher intra-population diversity for genomic DNA is expected (if for no other reason than the redundancy of the genetic code) and it is also expected that different genetic elements (for example, chromosome inversions, allozymes, DNA haplotypes and nucleotide sequence varia-

tion) with increasing power of resolution will show progressively higher levels of geographical differentiation. This is true of all species.

Rama S. Singh

*Department of Biology,
McMaster University,
Hamilton, Ontario L8S 4K1,
Canada*

Larry R. Hale

*Department of Biology,
University of Prince Edward Island,
Charlottetown,
Prince Edward Island C1A 4P3,
Canada*

1. Begun, D. J. & Aquadro, C. F. *Nature* **365**, 548–550 (1993).
2. Hale, L. R. & Singh, R. S. *Proc. natn. Acad. Sci. U. S. A.* **83**, 8813–8817 (1986).
3. Hale, L. R. & Singh, R. S. *Molec. Biol. Evol.* **4**, 622–637 (1987).
4. Hale, L. R. & Singh, R. S. *Genetics* **129**, 103–117 (1991).

BEGUN AND AQUADRO REPLY — A key difference between our study of nuclear genes and that of Hale and Singh on mtDNA is that the genetic history and potential of these *Drosophila* populations (in the broadest sense) is to be found in their nuclear DNA (the information content of mtDNA is minimal).

The qualitative trends of more variation in Africa and geographical differentiation between Africa and US populations were seen both in Hale and Singh's data and in ours. However, our data are from a very large number of genetically independent nuclear gene polymorphisms, whereas the mtDNA data are from a single, non-recombining, maternally inherited marker. Therefore, one should have far more confidence in drawing general conclusions from our data.

Hale and Singh's points about the comparison of mtDNA and protein electrophoretic data are not relevant to our conclusion that the heterozygosity in Zimbabwe is greater than in the United States, based on comparison of an equivalent set of nuclear DNA markers. Although mtDNA may show greater geographical differentiation than nuclear genes (as a result of the smaller effective size), there is no reason to expect that protein variation and nuclear restriction site polymorphism should show different degrees of population differentiation (assuming that the observed variation has no fitness consequences). It is highly unlikely that the contrast between the level of protein polymorphism in West Africa and the level of DNA heterozygosity in Zimbabwe is simply explained by the higher resolution of restriction site analysis.

David J. Begun

Charles F. Aquadro

*Section of Genetics and Development,
Cornell University,
Ithaca, New York 14853,
USA*

1. Shepard, R. N. *Psychometrika* **27**, 125–140 (1962).
2. Shepard, R. N. *Psychometrika* **27**, 219–246 (1962).
3. Kruskal, J. B. *Psychometrika* **29**, 1–27 (1964).
4. Kruskal, J. B. *Psychometrika* **29**, 115–129 (1964).
5. Young, M. P. *Nature* **358**, 152–155 (1992).
6. de Leeuw, J. & Bettonvil, B. *Psychometrika* **51**, 149–153 (1986).
7. Spence, I. & Ogilvie, J. C. *Multivariate Behav. Res.* **8**, 511–517 (1973).
8. Wilkinson, E. M. in *Mathematics in the Archaeological and Historical Sciences* (eds Hodson, F. R., Kendall, D. G. & Tautu) 276–283 (Edinburgh University Press, 1971).
9. Hubert, L. Bri. J. *Math. Stat. Psychol.* **27**, 133–153 (1974).
10. Young, M. P. *Proc. Roy. Soc. B*, **252** 13–18 (1993).
11. Scannell, J. W. & Young, M. P. *Curr. Biol.* **3**, 191–200 (1993).
12. Young, M. P., Scannell, J. W., Burns, G. A. P. C. & Blakemore, C. *Rev. Neurosci.* (in the press).
13. Gower, J. C. in *Mathematics in the Archaeological and Historical Sciences* (eds Hodson, F. R. *et al.*) 138–149 (Edinburgh University Press, 1971).
14. Stenson, H. H. & Knoll, R. L. *Psych. Bull.* **71**, 122–126 (1969).
15. Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (eds Ingle, D. G. *et al.*) 549–586 (MIT, Cambridge, 1982).
16. Felleman, D. J. & Van Essen, D. C. *Cerebral Cortex* **1**, 1–47 (1991).
17. Zeki, S. & Shipp, S. *Nature* **335**, 311–317 (1988).
18. Morel, A. & Bullier, J. *Visual Neurosci.* **4**, 555–578 (1990).
19. Kendall, D. G. in *Mathematics in the Archaeological and Historical Sciences* (eds Hodson, F. R. *et al.*) 213–247 (Edinburgh University Press, 1971).
20. Scannell, J. W. *et al. Invest. Ophthalmol. Vis. Sci.* **35**, 3262 (1994).

A penchant for protists

Laurence D. Hurst

Tracing the History of Eukaryotic Cells: The Enigmatic Smile. By Betsey Dexter Dyer and Robert Alan Obar. Columbia University Press: 1994. Pp. 259. \$30, £20 (pbk); \$65, £43 (hbk).

THE vast menagerie of early eukaryotes that constitute the Protista is as bizarre an array of the weird and wonderful as can be imagined: from organisms without sex but which nevertheless alternate ploidy, to others that are parasitized by their sister species and almost everything in between. While God, as J. B. S. Haldane once ventured, must have "an inordinate fondness for beetles", I suspect that the Deity has also, at the very least, a penchant for protists.

The mind-boggling assortment is not just a curiosity. In the details of protist biology may also lie clues to the origins and evolution of eukaryotes. Just as the Cheshire cat vanished leaving only its smile, so too history leaves a scar on present-day descendants. It is these enigmatic fragments that Dyer and Obar seek to tease out.

Names, dates, births and marriages are as important in understanding eukaryotic history as they are for more conventional history. For the events that the authors discuss, the fossil record is useful for estimating dates, whereas molecular data are making significant inroads into reconstructing the family tree of unicellular life. Unfortunately, as the authors note, data from a few key eukaryotic groups are missing and, as is almost inevitable, there is disagreement over the phylogeny. This problem is compounded by a turbulent nomenclature (a turbulence typified, for instance, by this reviewer's frequent conflation of protist with protoctist). But Dyer and Obar treat all the terminological entanglement pragmatically and do not let the muddle obscure the information.

For the history of the eukaryotes, however, it is the marriages that are most noteworthy. Although only 25 years ago it was considered unorthodox and largely vilified, the idea that most eukaryotic cells are a union between symbiont and host is now almost unquestioned orthodoxy, at least for the ancestry of plastids and mitochondria. Today's problems, well reviewed by the authors, are numbering the independent origins and deciding which organelles are symbiotically derived and from which taxa.

Although the symbiogenic hypothesis had old roots, the acceptance of the idea is in large part due to the advocacy of Lynn Margulis. Some of her zeal has transferred to her students, Dyer and Obar. Although they would clearly be delighted if certain organelles could be shown to have sym-

biotic origins, they usually balance their enthusiasm with words of caution and present the alternative arguments. A useful discussion of the types of data that can be used to examine the symbiogenic hypothesis is provided. They leave, however, a nagging loose end, the issue of whether in principle the origin of most organelles can ever be resolved.

Consider for instance the authors' new hypothesis for a symbiotic derivation of microtubule organizing centres (MTOC). By what information will the hypothesis live or die? Dyer and Obar argue, quite correctly, that an absence of MTOC-associated DNA would be evidence neither for nor against a symbiogenic origin: it may never have been there or it could have been transferred to the nucleus. So what is good evidence? That some prokaryotes may have proteins resembling those in the MTOC is not itself conclusive. It may be explained as convergence, the product of symbiosis, of horizontal gene transfer, or may simply reflect nuclear genome ancestry. The reason we can be fairly certain of mitochondrial and plastid origin is that so many facts point to the same end. For organelles without DNA the ancestry might, I regret, evade a convincing resolution.

For every history, names, dates and family trees are only part of the story. There is also the narrative, the enigmatic 'why?'. Why, for instance, did sex and sexes evolve and why were symbiotic relationships so successful? Testing explanations of some of these problems may be as hard as resolving issues of organelle origin. This is particularly true for events that happened only once. Suggestive patterns do however exist. Many of the major events in prokaryote and eukaryote evolution are, for instance, coincident with changes in atmospheric content. Although the authors provide stimulating speculation as to what this might mean, we are left with possibilities, maybes and,

quite often, no clear means to resolve the uncertainties.

The prospect is, however, less bleak for phenomena that have multiple evolutions. With an adequate phylogeny, independent evolutionary events can be resolved and phylogenetic contrasts sought. If, instead of apparently meaningless diversity, we are to have patterns, then this, the comparative method, seems the best analysis available. Some discussion of the methodology would then have been welcome, as would mention of attempts to perform such analysis. In several instances, no doubt for want of space, the underlying evolutionary genetic models also go unacknowledged. I can, however, only concur with the authors' plea for more work on protist biology. I must also concur with their admiration for the most recent compilation of knowledge, the remarkable *Handbook of Protoctista* edited by Margulis *et al.* (Jones and Bartlett, 1991), possibly her second great contribution to science. With an updated copy of this in one hand and a well resolved phylogeny in the other, we may soon have answers to many of the burning questions not only of early eukaryote history but also of evolution in general. □

Laurence D. Hurst is in the Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.



PART of a temple scroll showing four Daoist genii. From the cover of *The Shorter Science and Civilization in China: 4*, an abridgement by Colin A. Ronan of Joseph Needham's monumental work. The fourth volume covers advances in mechanical engineering in early and mediaeval China. Cambridge University Press, £40, \$19.95 (hbk); £19.95, \$34.95 (pbk).

Is chemistry physics?

Kostas Gavroglu

From Chemical Philosophy to Theoretical Chemistry: Dynamics of Matter and Dynamics of Disciplines, 1800–1950.

By Mary Joe Nye. *University of California Press*: 1994. Pp. 328. \$48.

IN 1929, Paul Dirac declared that “the underlying physical laws for the mathematical theory of a large part of physics and the whole of chemistry are completely known”. It seemed that the only thing left for chemists to do was to devise methods for carrying out the calculations. His optimism is reminiscent of the ‘grand scheme’ proposed by Laplace at the end of the eighteenth century or, more recently, talk about ‘theories of everything’. In these pronouncements there is a touch of arrogance, a product, no doubt, of the phenomenal success of quantum mechanics (Dirac), Newtonian mechanics (Laplace) or gauge theories. But arrogance in science, just as in politics, tends to obscure the historical nature of problems. The statements adopt a strongly reductionist view of the world, implying that, immense mathematical difficulties notwithstanding, the variety and complexity of entities and phenomena are constructible from and hence explainable in terms of ‘basic entities’ and underlying laws. Dirac’s pronouncement had a further, rather paralysing, implication: that chemistry, in the final analysis, is physics. It was not the first time that chemistry was portrayed as a branch of physics.

Mary Joe Nye attempts to trace the history of this ambivalent attitude towards chemistry during the past two centuries. The history of chemistry has been marked by a Sisyphean curse: the more chemistry resembled physics, the more it was regarded as being a ‘true’ science; but the more it gained such prestige, the more it lost its identity, and the less autonomous it became. Nye directly addresses the question of whether chemistry is reducible to physics, but she also examines the issue by studying the formation of the boundaries of physical chemistry, chemical physics, theoretical chemistry and quantum chemistry. She convincingly demonstrates that the construction of the identity of theoretical chemistry and the drawing of disciplinary boundaries were due to a network of factors. She examines genealogy and historical mythology at the beginning of each discipline; archetypal language and imagery; codified practices and rituals; the establishment of institutions based on rights and responsibilities; and external recognition, shared values and unsolved problems. Before turning to

specific case studies, she discusses chemical epistemology, and language and imagery in nineteenth-century chemistry.

Nye is committed to the view that research schools are the most convenient units for the study of discipline formation, so she concentrates on the Paris and London–Manchester schools of theoretical organic chemistry between 1880 and 1930, on Christopher Ingold’s attempt to integrate physical and organic chemistry and on the development of quantum chemistry in Britain and America. She steers readers through a maze of details and arguments to conclude that theoretical chemistry slowly acquired its autonomy as a result of a complex interplay of the convergent and divergent practices of individuals and schools. The establishment of physical chemistry and quantum chemistry was due not merely to the successful application of physical methods and theories to chemical problems, but also to a series of legitimizing procedures.

Two aspects in specifying the identity of disciplines deserve separate consideration; their significance is, I think, played

down in Nye’s treatment. The first is the role of disputes, which can reveal subtle undercurrents in the processes of legitimizing new theories, defining boundaries and inaugurating a particular ethos for each discipline. The second aspect is the meaning of theory in chemistry, a discipline in which empirically determined parameters are an integral part of theory. Both aspects involve the reappraisal of the praxis of chemists, which reflects the commitments and methodological preferences of the community as a whole.

Nye has had to be very selective in her choice of case studies. Other choices might even have provided counter-arguments to parts of her thesis. Nevertheless, the book is an intriguing attempt to contribute to the understanding of an extremely important and difficult theoretical problem in the history of the physical sciences, and, all in all, it is a success. □

Kostas Gavroglu is a historian in the Department of Physics, Zografu Campus, National Technical University, 15780 Athens, Greece.



STAR shots: top, Chile’s Southern Observatory; bottom left, the 305-meter Arecibo radiotelescope in Puerto Rico, which began searching for extraterrestrial life in 1992; right, astronomer Gerry Luppino eyes up a Hubble Space Telescope unit of 640,000 cells. The pictures won Roger H. Ressmeyer of *National Geographic* magazine first prize in the Science and Technology category of the 37th World Press Photo Contest held in 1993. Taken from *World Press Photo Yearbook 1994* edited by Ben Ten Berge and Keri Lundelin. Thames and Hudson, £9.95 (pbk).

Against method?

Janet Sayers

A Most Dangerous Method: The Story of Jung, Freud and Sabina Spielrein. By John Kerr. Sinclair-Stevenson/Knopf: 1994. Pp. 607. £25.

THIS is not a love story, John Kerr warns. Instead he portrays the romance (1907–1913) between Freud, Jung and Sabina Spielrein as a passionately engaging but ultimately “gruesome ghost story” of science betrayed.

Kerr, a clinical psychologist, sets the stage with an account of nineteenth-century scientific materialism. In Zurich, Jung’s word-association experiments were promising to substantiate the otherwise immaterial ‘complexes’ underlying mental illness, with many of his patients, including Spielrein, showing delayed reaction times to particular stimulus words. Meanwhile Freud, in Vienna, was bothered by his patients’ transference onto him of their erotic wishes and desires. He was also trying to re-explain sexual repression; his break with his former colleague Fliess had stopped him attributing it to bisexuality. It was hoped that Freud would soon write a handbook so others could use his method independently, allowing them reliably to interpret and test their patients’ free associations in terms of his theories without risk of contamination by suggestion.

But no such manual appeared. Instead, psychoanalytic attention switched from science to symbolism. Freud turned to analysing Jensen’s novel *Gradiva*, the symbols in Leonardo’s paintings and the myth of Oedipus, in an attempt to discover the key to the single ‘core’ sexual complex causing neurosis.

No harm in symbols. They have often been used to advance science. (Kekulé’s benzene ring, for example, emerged from his dream of snakes biting each other’s tails.) True, as Spielrein pointed out on becoming a medical researcher herself, in madness symbols can become ossifying impersonal *idées fixes*. For her, however, the symbol of Wagner’s Siegfried (misinterpreted by Jung, she said, as her wish to bear his son) served to transform her otherwise backward-looking symptoms (including an obsession with thinking about faeces as she ate, and about her father spanking her brother) into an ideal that she could project herself into in the future.

As for Freud and Jung, their prospects of scientific collaboration ended, Kerr argues, not as others have claimed from the revelation of Jung’s sexual liaison with Spielrein (their “poetry” as she put it), but from her scientific papers to the Vienna Psychoanalytic Society. These papers indicated Jung’s rejection of Freud’s sexual

theory of neurosis. She herself went on to develop the hypothesis that sexual repression is due to the threat of self-annihilation involved in immersing oneself sexually in someone else.

Her mentors, however, turned to explaining mental illness in terms of a mythic past. Jung had previously regarded mythic symbols as a means of helpfully recasting and thereby better understanding the incestuous regressions triggered in patients by adolescent conflict. Now he regarded symbols as the cause of both health and disease. He himself nearly succumbed. He became so immersed in an image of a



gorgon mother trapping her hero son in the underworld that he could barely surface to complete an article on symbols and their transformative effects.

Freud too plunged into prehistory. Neurosis, he now insisted, stems from the killing of the father by the primal horde — that this is the source of the Oedipus complex. No need, therefore, to look any further for proof of the complex’s universality or for validation of its repression as the cause of a patient’s symptoms. Nor any need, accordingly, for a primer on psychoanalytic method.

So all hope of psychoanalysis as science died, Kerr claims, apart from a brief flickering with Jung’s theory of psychological types — of the extrovert as prone to the transference of neurosis and of the introvert as prey to the inward-looking preoccupations of psychosis. At least these character traits could, and indeed have been, empirically tested.

Kerr’s tale finishes, however, with Freud misrecording Spielrein’s theory of sexual repression as a precursor of his theory of the death instinct; with Spielrein herself abandoning psychoanalysis first for musical composition and then for work with the psychologists Piaget, Vygotsky

and Luria; and with Jung using her as muse for his theory of the anima, which, at the end of his life, he carved into a stone triptych.

As for psychoanalysis, devoid of a clearly stated method for investigating and testing its theories, it became at best a matter of creative guesswork, at worst a matter of indoctrination. No wonder, therefore, that its organizations and journals, which Freud and Jung had done so much to found, quickly deteriorated from platforms for the free exchange of scientific ideas into quasi-religious institutions devoted not to discovering the truth but to propagating and perpetuating dogma, not least through insinuating tittle-tattle against any who dared to dissent.

It is a sorry tale. But to the bitter end Kerr keeps us enthralled with his exhilaratingly informative, sympathetic and thought-provoking account of the real insights and genius that went into the original making of psychoanalysis. He thereby admirably reverses, and in the process explains, the recent trend towards recounting its history not as science but as gossip. □

Janet Sayers is in Keynes College, University of Kent, Canterbury, Kent CT2 7NP, UK.

Processes and origins

J. R. L. Allen

Geomorphology of Desert Environments. Edited by Athol D. Abrahams and Anthony J. Parsons. Chapman and Hall: 1994. Pp. 674. £85, \$16.95.

DESERT environments can support only a limited range and density of specialized organisms, but they unquestionably present a wide and rich variety of geomorphological challenges, if the thickness, weight and length of this book are any criteria.

Geomorphology of Desert Environments contains 26 review chapters prepared by 22 international experts and other active workers from the United States (13), the United Kingdom (5) and Australia, Canada, Germany and New Zealand (1 each). The grouping of chapters into eight sections helps to give a comprehensive account of the geomorphology of deserts — that is, areas of sub-humid to hyper-arid climate — with special emphasis on operative processes.

Part one (two introductory chapters) deals with desert environments generally and presents a geomorphic comparison of the main desert regions. Weathering in deserts is covered in part two (four chapters) dealing with weathering processes and forms, aridic soils, patterned ground and desert pavements, duricrusts and rock

varnish. Part three (hillslopes) covers rock slopes, rock-mantled slopes and badlands, and includes a brief but useful chapter by C. Francis on vegetation.

The two chapters in part four (rivers) cover catchment and channel hydrology, and channel processes and forms. Part six introduces the hydrographic and geomorphic patterns of semi-arid lake basins, a subject typically played down or even ignored in most books on desert geomorphology. Part seven (aeolian surfaces) covers aeolian sediment transport, dune morphology and dynamics, and landforms resulting from aeolian erosion. The last seven chapters, grouped under the title "Climatic Change", bring together many of the themes treated earlier, but in the important context of the role of time and climate variation in the development of today's desert geomorphology, after the operation of the extreme and rapid climate changes of the later Cenozoic era and the Quaternary period.

In addition to the scientific challenges, the sensitivity and human importance of desert environments, and especially the vexed issue of desertification, make it unsurprising that books on desert geomorphology and processes are published so often. Happily, *Geomorphology of Desert Environments* is a worthy and desirable newcomer, amplifying and in many respects complementing *Desert Geomorphology* by R. U. Cooke, A. Warren and A. S. Goudie (UCL Press, 1993; for a review see *Nature* **363**, 313; 1993).

The new work's emphasis on North America is not objectionable and there is little to disappoint. Generally, the chapters are comprehensive and thoroughly researched, with lengthy bibliographies, and several contain original material. The insights into the origins of badlands and hillslopes that A. D. Howard has gained using simulation models fully justify the application of the technique and suggest that its use would be beneficial in other geomorphological realms. D. R. Currey gives an incisive hydrographic and geomorphological analysis of semi-arid lake basins and their sediment fills — perhaps limitations of space prevented him from further illustrating his case with field photographs and maps.

The book is generally well illustrated, but the reader may find that some of the photographs are too small and their contrast excessive. Some of the line drawings are fuzzy and poorly lettered. □

J. R. L. Allen is in the Postgraduate Research Institute for Sedimentology, University of Reading, Whiteknights, PO Box 227, Reading RG6 2AB, UK.

■ *Geomorphology in the Tropics: A Study of Weathering and Denudation in Low Latitudes* by Michael F. Thomas has just been published by Wiley at £85.

From DNA to databases

Chris Rawlings

Guide to Human Genome Computing.

Edited by Martin J. Bishop. *Academic Press*: 1994. Pp. 350. £25, \$39.95 (pbk).

Computer Analysis of Sequence Data.

Two volumes. Edited by Annette M. Griffin and Hugh G. Griffin. *Humana*: 1994. Pp. 372 and 433. \$59.95 each.

MOST molecular biologists and geneticists today rely on computer software for collecting data, retrieving sequence and genetic information from databases, or interpreting sequence or recombination frequency data. But programs remain complex and difficult to master. The mismatch between the software skills and the scientific needs of molecular biologists not only reflects the pace of new technological development; it also reveals a weakness in education and training. The management and analysis of data could well prove to be a bottleneck in genome projects.

So we should welcome these two publications, which aim to make software more accessible and scientists more aware. Although ostensibly covering the same topic areas, the books adopt different approaches. *Guide to Human Genome Computing* is more conventional, each chapter providing an introduction to the principles of software together with guidance on the use of programs, and liberally illustrated with examples and figures. The coverage is broad and aimed at the molecular geneticist involved in a variety of research areas from genetic linkage analysis, physical mapping data management, sequence assembly, sequence comparison, inferring biochemical function from sequence data and the use of comparative genetic maps. Many authors go to great lengths to compare a range of different software tools, showing that there is generally no best solution to complex problems such as genetic linkage and sequence comparison — different programs give different answers depending on how they are used.

Much of the material in this book arose from the training courses run at the UK Medical Research Council Human Genome Project Resource Centre. Occasionally, some parochialism shows through. But overall the book will appeal to a wide community of genome scientists.

Computer Analysis of Sequence Data comes in two volumes and focuses on programs rather than problems. The main sections cover the Genetics Computer Group (13 chapters) and the Staden (13 chapters) software suites, software for IBM-compatible personal computers (in-

cluding the PC/Gene programs (6 chapters) and also, rather surprisingly, the now obsolete MicroGenie software (5 chapters)) and programs for the Apple Macintosh (including MacVector (5 chapters) and DNA Strider). Other chapters deal with individual programs, and include a useful overview of methods for multiple protein sequence alignment while introducing some of the problems of phylogenetic inference.

The programs are presented in a practical step-by-step fashion, and the notes at the end of most chapters provide useful hints and warnings. But despite some excellent contributions, the value of the book as a whole is marred by a lack of illustrations and by repetition resulting from the rigid "Materials-Methods-Results" structure. Also, five chapters appear in both volumes (presumably some people will buy only one or the other). Although the choice of repeated material is for the most part reasonable (covering sequence comparison with the FASTA program, the submission of sequence data to the database curators and the conversion of data to different data formats), the duplicated chapter about Internet is somewhat dated and parochial. There are other signs that the book was conceived several years ago: for example, it is surprising that there is no mention of the BLAST program for rapid sequence database searching. Finally, the rather inadequate index contains no cross-references between the two volumes.

Nevertheless, the detailed program recipes in *Computer Analysis of Sequence Data* will provide a useful adjunct to user manuals. □

Chris Rawlings is in the Biomedical Informatics Unit, Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK.

New in paperback

Collins Dictionary of Astronomy

edited by Valerie Illingworth. A new edition, revised and updated, of the *Macmillan Dictionary of Astronomy*, which was first published in 1979 and revised for a second edition in 1985. A team of 24 professional astronomers contribute more than 3,500 entries. HarperCollins, £8.99.

In the Beginning: The Birth of the Living Universe

by John Gribbin. A guide for the lay person to ideas about the connection between cosmology and evolution. Penguin, £6.99.

Defenders of the Text: The Traditions of Scholarship in an Age of Science, 1450

— 1800 by Anthony Grafton. A series of scholarly but lively essays on the relation between humanism and science, focusing on Poliziano, Scaliger, Kepler and Wolf. Harvard University Press, \$20.25, £13.50.

Crystal structure of human chorionic gonadotropin

A. J. Laphorn, D. C. Harris, A. Littlejohn, J. W. Lustbader*,
R. E. Canfield*, K. J. Machin†, F. J. Morgan‡ & N. W. Isaacs§

Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK

* Department of Medicine, Columbia University, New York 10032, USA

† St Vincent's Institute of Medical Research, Melbourne, Victoria 3065, Australia

‡ Department of Biochemistry, LaTrobe University, Melbourne, Victoria 3083, Australia

The three-dimensional structure of human chorionic gonadotropin shows that each of its two different subunits has a similar topology, with three disulphide bonds forming a cystine knot. This same folding motif is found in some protein growth factors. The heterodimer is stabilized by a segment of the β -subunit which wraps around the α -subunit and is covalently linked like a seat belt by the disulphide Cys 26–Cys 110. This extraordinary feature appears to be essential not only for the association of these heterodimers but also for receptor binding by the glycoprotein hormones.

In humans the early stages of pregnancy are maintained by the hormone chorionic gonadotropin (hCG) from the conceptus. This hormone is structurally related to the anterior pituitary gonadotropins, follicle-stimulating hormone (FSH) and luteinizing hormone (LH), which regulate the cellular and endocrine function of the ovary and testis. Together with thyroid-stimulating hormone (TSH) they constitute the family of glycoprotein hormones. These consist of two subunits, α and β , which associate non-covalently to form a heterodimer; in a given species the α -subunits are identical and the β -subunits are different (but homologous) for the different hormones. Although the heterodimer is required for receptor binding, it is the β -subunit that determines the specific activity of each hormone. The hormones are all glycosylated with *N*-linked complex carbohydrates, which confer heterogeneity on a given hormone; hCG is the most heavily glycosylated with four additional *O*-linked carbohydrates on the serine-rich C-terminal extension of the β -subunit. In human glycoprotein hormones, the common α -subunit contains 92 amino acids, with 10 half-cystine residues which form five intramolecular disulphide linkages. The β -subunits vary in size from 114 residues in LH to 145 in CG and contain 12 half-cystines which form six conserved disulphide bridges. There is a high degree of sequence similarity in the first 114 amino acids between hCG and the other hormones (LH 85%; FSH 36%; TSH 46%). The homology between hCG and LH reflects a common biological function, as both proteins bind the same receptor; FSH and TSH bind to structurally similar but distinct receptors.

Here we present the crystal structure of hCG, report the correct disulphide pairings in both subunits, and describe the overall protein fold and the dimer association that involves a structurally unique 'seat-belt' arrangement. The receptor-binding site is described. We show that both subunits of glycoprotein hormones are members of a structural superfamily of cystine-knot growth factors.

Structure determination and refinement

It was previously thought that heterogeneity of carbohydrate prevented any of the glycoprotein hormones from crystallizing. The bulk of the carbohydrate can be removed from hCG by treatment with anhydrous hydrofluoric acid, leaving the deglycosylated protein with the ability still to bind receptors,

although post-translational effects are not triggered¹. Circular dichroism shows that the structures of the deglycosylated and native hormones are identical². Using the crystallization conditions for HF-treated hCG (HF-hCG)³, isomorphous crystals of neuraminidase-treated hCG have been grown in which only the terminal sialic acid residues have been removed⁴.

HF-treated hCG crystallizes from ammonium sulphate solutions³ as hexagonal bipyramids of space group *P*6₃22 and cell dimensions $a = 88.68$, $c = 177.24$ Å. The crystals diffract to 3.5 Å resolution on a laboratory X-ray source and to 2.5 Å with synchrotron radiation, but are very sensitive to radiation damage. The structure has been determined to 3.0 Å resolution by multiple isomorphous replacement (MIR) and maximum entropy (ME) solvent flattening; Table 1 gives the details and statistics for the phasing procedure (details of the ME phasing procedure will be published elsewhere).

The final electron density map allows a tracing of the α -subunit from residues 5 to 89 and the β -subunit from 2 to 111. The residue β 111 is open to a large 50 Å diameter solvent channel and it is assumed that the remaining 34 C-terminal residues adopt a random conformation and are therefore not visible in the map.

The initial model was refined using the program XPLOR⁵. Five rounds of refinement and model building gave a final structure with a crystallographic *R* factor of 21.8% for data from 10–3 Å. No solvent molecules, but six *N*-acetylglucosamine molecules positioned in good density have been included. A geometric analysis of the refined structure with the program PROCHECK⁶ gave values better than expected for structures determined at this resolution.

The disulphide bridges

There are five disulphide bonds in the α -subunit and six in the β -subunit. The chemical assignment of the disulphide pairings has been extensively studied (reviewed in ref. 2). The assignments of Mise and Bahl^{7,8} are considered the most reliable, with α 7–31, α 10–32, β 93–100, β 26–110 and perhaps β 23–72 assumed to be correct. From the ME phased map, some electron density could be interpreted as disulphide linkages 23–72, 26–110 and 34–88 of the β -subunit. The remaining S–S bridges could not be formed as expected (that is, between residues 38–57, 9–90) and the map could be interpreted only if the linkages were from residues 9 to 57 and 38 to 90, thereby forming an unusual knot

§ To whom correspondence should be addressed.

of three disulphide bridges. The density for the α -subunit contained a similar cystine knot, and for the amino-acid sequence to fit, disulphide linkages 7–31, 10–60, 28–82, 32–84 and 59–87 were necessary.

This cystine-knot motif⁹ is found in three growth factors: NGF¹⁰, TGF- β ¹¹ and PDGF-BB¹². The motif involves three disulphide bridges arranged so that two disulphides link adjacent antiparallel strands of the peptide chain and form a ring penetrated by the third disulphide (Fig. 1). The disulphide pairings are: α (7–31, 10–60, 28–82, 32–84, 59–87) and β (9–57, 23–72, 26–110, 34–88, 38–90, 93–100) (Figs 1 and 2). Of the linkages not previously predicted (α 10–60, 28–82, 32–84; β 9–57, 38–90), all are involved in the cystine knots.

Structure of the protein subunits

Figure 2a shows that the subunits have similar folds determined largely by the central cystine knot. On one side of the knot there is a loop of double-stranded β -sheet-like structure (the long loop); on the other side there are two hairpin loops lying

in almost parallel planes. In the β -subunit, these hairpins are linked by the Cys 23–72 disulphide bond.

In the C-terminal hairpin loops of each subunit, there are a pair of β -sheet bulges at adjacent positions in each strand (α 62–65, 79–82; β 59–62, 85–88). In the β -subunit, the bulges have a classical hydrogen-bonding pattern; in the α -subunit the pattern is less complete. The structures of all the cystine-knot growth factors have β -sheet bulges located immediately after Cys(V), suggesting they are an integral component of this motif.

The long loop in the α -subunit consists of a stretch of antiparallel β -sheet made up of residues 35–39 and 52–56. The three residues Pro38–Thr39–Pro40, which are conserved in all 18 α -subunit sequences determined, break the β -sheet away at an angle of about 100° and lead into two turns of a 3_{10} -helix involving residues 40–46. Asn 52 is also on a bend that directs the N-linked carbohydrate on this residue out from the interface between the α - and β -chains.

The corresponding region of the β -subunit forms a very open loop with no main-chain hydrogen bonds between the antiparal-

TABLE 1 Summary of crystallographic analysis

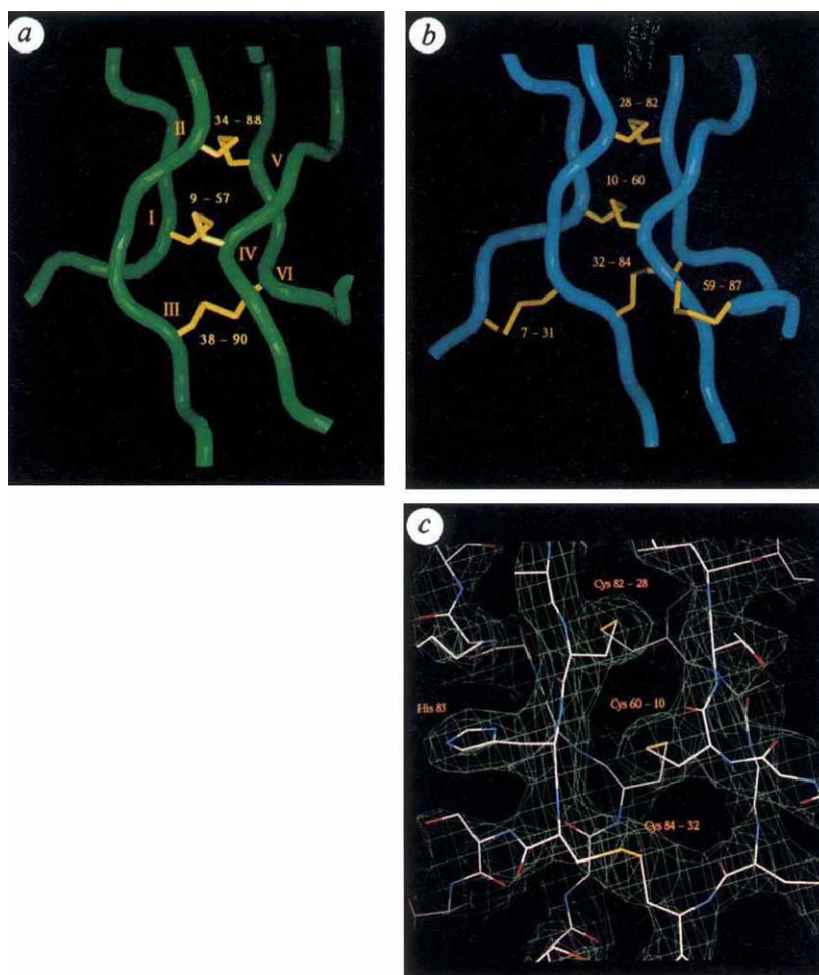
Diffraction data	Native	K ₂ Pt(CN) ₄	AgNO ₃	AgNO ₃	Hg(OOCH ₃) ₂	KAu(CN) ₄	K ₂ Pt(CN) ₄
Data set	Xen./Film	Xen.	Xen.	Xen.	Film	Film	Film
Detector	Xen./Film	Xen.	Xen.	Xen.	Film	Film	Film
Resolution (Å)	3.0	3.5	5.0	3.5	3.5	3.9	3.3
Unique reflections	8,365	3,169	1,096	3,664	2,473	2,272	2,949
Completeness (%)	94.9	58.9	81.6	70.6	44.0	55.2	52.4
R _{merge} (%)	11.1	6.0	7.7	7.5	10.0	9.1	7.5
MIR phasing							
Mean fractional ID		0.126	0.094	0.115	0.080	0.148	0.073
Number of sites		2	2	3	1	2	1
Rc		0.73	0.68	0.82	0.80	0.80	0.78
Phasing power	centric	1.1	1.0	0.7	0.7	0.7	0.6
	acentric	1.4	1.5	1.1	0.9	1.0	0.6
Phase combined FOM	0.52						

Crystallization and data collection: Crystals were grown at 25 °C as described³. More than 20 heavy-atom compounds were tested with a variety of soak conditions to yield four useful derivatives. Diffraction data were collected on film on station 7.2 ($\lambda = 1.48$ Å) at the SERC Synchrotron Radiation Source, Daresbury. Film data were processed using the MOSFLM program and native and derivative data were scaled using SCALEIT from the CCP4 suite of programs⁴⁰. Area detector data were collected on a Siemens/Xenotronics P1000 area detector, and processed using the XDS program⁴¹. The final native dataset used for refinement, on which the current model is based, is a combination of a complete Xenotronics dataset to 3.5 Å resolution collected from a single crystal, merged with the less complete film data collected from several crystals which extends to 2.8 Å. From a consideration of the merging *R* factor statistics, only data to 3.0 Å were used. XEN, Xenotronics area detector; $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i$, where I_i is the intensity of an individual reflection and $\langle I \rangle$ is the mean intensity of that reflection; ID, isomorphous differences.

MIR analysis: The Pt derivative was solved using SHELXS86⁴² and the other derivatives by a combination of difference Fourier syntheses and difference Patterson maps. There are essentially three sites; the major site shared by the Pt and Au derivatives is in a pocket on a 2-fold axis of symmetry binding β 45 Arg and its 2-fold symmetry equivalent. Both the primary Ag site and the only Hg site are very close (sandwiched between α 29 Met and β 41 Met) and the other Ag sites are at α 83 His and α 71 Met. The heavy-atom parameters were refined and MIR phases calculated using MLPHARE for the six derivative datasets, using both isomorphous and anomalous differences to give a FOM of 0.52. The correct space group was determined from refinement of heavy-atom sites in MLPHARE using anomalous data from the Pt derivative and confirmed by the handedness in the twist of β -sheets in the final structure. Cullis *R*-factor, $R_c = \sum ||F_{\text{PH}} \pm F_{\text{P}} - F_{\text{H}}| / \sum |F_{\text{PH}} - F_{\text{P}}|$ where F_{P} and F_{PH} are the structure factor amplitudes for native and derivative data respectively and F_{H} is the calculated heavy-atom structure factor; phasing power = $\langle F_{\text{H}} \rangle / \langle E \rangle$, where $\langle F_{\text{H}} \rangle$ is the r.m.s. of the heavy-atom scattering factor and $\langle E \rangle$ is the r.m.s. lack of closure; the summation is computed for reflections used in the heavy-atom refinement cycle. FOM, figure of merit.

Density modification and model building: The initial phases were refined at 3.5 Å resolution using the Wang solvent-flattening procedure⁴³. A protocol of five cycles of solvent flattening following an envelope redetermination (55–60% solvent) was used for 10 rounds of phase refinement. Phase extension from 4.0 Å using very small increments in resolution was used in an attempt to improve the quality of the maps. Although ~75% of the protein chain, including the major β -strands in the structure, could be traced, the maps were not sufficiently clear to define the subunit interface or to get a sequence alignment. Further density modification was achieved using the program MICE⁴⁴. Using MIR-phased reflections with FOM ≥ 0.70 as basis-set reflections and all data to 3.0 Å resolution, maximum entropy solvent flattening (MESF) was performed. Seven rounds of MESF were calculated with phase recombination computed using SIGMAA⁴⁵ and envelope determination using the Wang procedure. A round of phase permutation using 7 important unphased reflections gave phases for 5 of these, which were then combined with the initial MIR basis-set reflections and an additional seven rounds of MESF computed. Partial structure phasing from 14 peptide fragments, with side chains assigned according to size was included in phase recombination in the later rounds of MESF. This led to a map which was sufficiently unambiguous to allow the complete tracing of both subunits. **Refinement:** The polypeptide chain was built into a combined map using the programs FRODO⁴⁶ and O⁴⁷. This model was refined at 3.5 Å using the program XPLOR⁵. Five rounds of model rebuilding and refinement using both conventional positional refinement and simulated annealing, during which the resolution was extended to 3.0 Å, resulted in the current model. The model comprises a total of 1,565 non-hydrogen atoms. Temperature factors were refined individually and then averaged by group with a final overall *B* value of 26.5 Å². Data from 10 to 3.0 Å with $F > 2\sigma(F)$ (7,877 reflections) were used in the final refinement to give an *R* factor of 21.8%, and a Free *R* factor (791 reflections) of 31.4%. The stereochemistry is typified by an r.m.s. bond deviation from ideality of 0.017 Å and r.m.s. angles of 2.28°. All but one of the non-glycine ψ and ϕ angles lie in allowed regions in the Ramachandran plot, with 77% in the most favoured regions. Atomic coordinates will be deposited in the Brookhaven Protein Data Bank.

FIG. 1 The structure of the α - and β -subunit cystine knots shows the common fold that comprises three disulphide bridges, arranged so that two disulphides link adjacent parallel strands of the peptide chain and form a ring through which the third disulphide penetrates. **a**, The common arrangement of the half cystines in the sequence⁹ is shown for the β -subunit. If the half cystines within the knot are assigned roman numerals to indicate their order in the protein sequence, disulphide bonds Cys(II)–Cys(V) and Cys(III)–Cys(VI) form the ring through which the interpenetrating disulphide Cys(I)–Cys(IV) is formed. There are two general sequence requirements, Cys(II)–X–Gly–X–Cys(III) and Cys(V)–X–Cys(VI), although this is not absolute as in NGF, where the central Gly in the first motif is replaced by a 7-amino-acid loop (Fig. 5). These sequence patterns are found in each subunit of hCG: α 28–32 (CMGCC), α 82–84 (CHC), β 34–38 (CAGYC) and β 88–90 (CQC). In the isolated subunits, the disulphides in the cystine knots show varying degrees of solvent accessibility, as predicted⁴⁸. Cys(I)–(IV) and Cys(II)–Cys(V) are buried in the knot, but Cys(III)–Cys(VI) is exposed to the solvent like the remaining disulphide bridges. **b**, In the α -subunit there are additional disulphide bridges (7–31 and 59–87) in close proximity to the knot. These invariant residues seem to play a role in formation of the heterodimer. There is a notable similarity in overall structure in the knots and the only significant differences, such as the dihedral angle of Cys(III)–Cys(VI), are due to conformational restraints imposed to accommodate the additional cystines in α . **c**, The $(2|F_{obs}| - |F_{calc}|)$ electron density map around the cystine knot of the α -subunit is indicative of the quality of the final map. The refined structure of the protein is illustrated as an atomic stick figure, colour coded for atom type: C, white; O, red; N, blue; S, yellow. The electron density is contoured at a 1σ level.



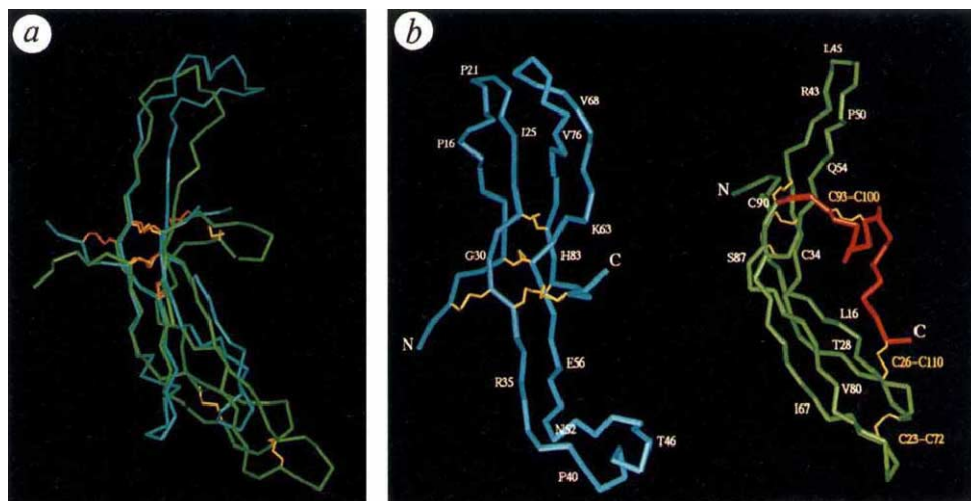
1el strands. Instead, the loop is stabilized by side chain-to-main chain hydrogen-bond interactions from Arg 43 to Pro 50 and Leu 52 and from Gln 54 to Met 41. One side of the loop, residues 38 to 45, is held by association with the α -subunit. In the crystal, a two-fold axis of symmetry causes the hydrophobic residues

48–50 to pack against the same residues from a neighbouring molecule.

Apart from the cystine-knot motif, there is no sequence similarity between the subunits, but when the cystine knots are superimposed, the similarity in structure is remarkable (Fig. 2b). The

FIG. 2 The structure of each subunit is shown as a $C\alpha$ trace with disulphide bridges. (The α -subunit is blue and the β -subunit is green in this and all subsequent figures.) Each molecule has a similar fold consisting of two β -hairpin loops on one side of a central cystine knot and a long loop on the other. **a**, The hairpin structure of each subunit is stabilized by a hydrophobic core extending between the two loops. The α -subunit core consists of α 17 Phe, α 18 Phe, α 25 Ile, α 68 Val, α 70 Val, α 71 Met, α 74 Phe and α 76 Val, with the three phenylalanines clustered together and partially exposed to the surface. The β -subunit core includes the residues β 16 Leu, β 18 Val, β 27 Ile, β 29 Val, β 67 Ile, β 69 Leu, β 80 Val and β 82 Tyr, which partly buries the disulphide bridge β 23–72.

The two hairpins of α are essentially regular, apart from an unusual bend at the invariant Asn 15 where the side chain hydrogen-bonds to the main chain NH of residues 17 and 18, resulting in Pro 16, Phe 17 and Phe 18 bulging away from the antiparallel strand and forming a prominent surface. Residues 90–111 are highlighted in red on the β -subunit. This 'seat-belt' region is a novel structure essential for forming



the heterodimer. **b**, Superimposing the α - and β -subunits at the cystine knot reveals the extent of their common structure. The $C\alpha$ position of 22 residues extending from the knot can be superimposed with a r.m.s. difference of 0.52 Å. Subunits are coloured as in **a**, with α -subunit disulphides in orange.

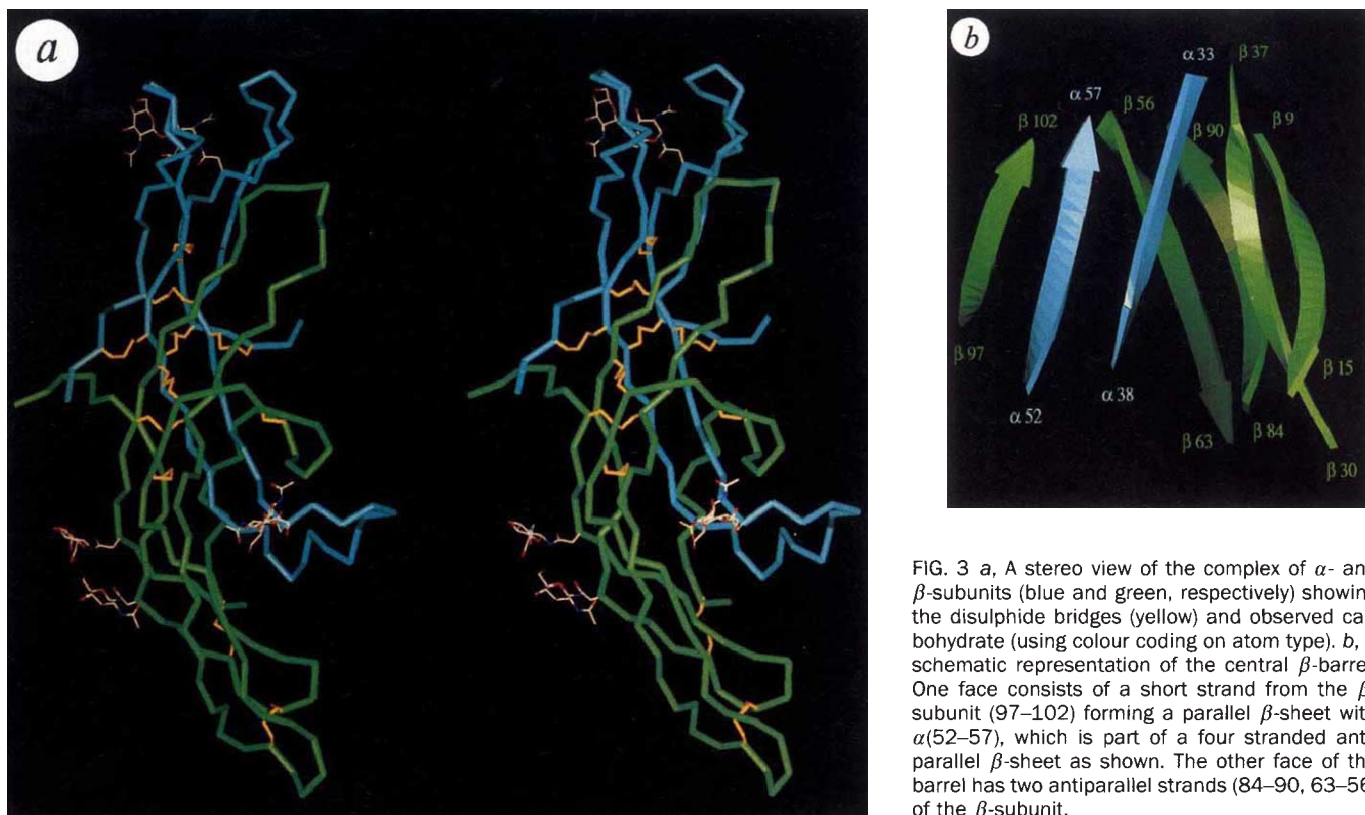


FIG. 3 *a*, A stereo view of the complex of α - and β -subunits (blue and green, respectively) showing the disulphide bridges (yellow) and observed carbohydrate (using colour coding on atom type). *b*, A schematic representation of the central β -barrel. One face consists of a short strand from the β -subunit (97–102) forming a parallel β -sheet with α (52–57), which is part of a four stranded antiparallel β -sheet as shown. The other face of the barrel has two antiparallel strands (84–90, 63–56) of the β -subunit.

two molecules have a common core structure and differ mainly in the extent of the loop regions. In the β -subunit the pair of hairpin loops are extended in size, while the long loop is larger in the α -subunit.

Formation of the heterodimer

Figure 3*a* shows the heterodimer. The two subunits are related by an approximate two-fold axis. Segments of well defined β -sheet structure near the cystine knot in each subunit are brought together to form a short seven-stranded β -barrel (Fig. 3*b*). The β 34–36 (CAGY sequence), long implicated in the formation of the dimer², is in the central strand. Away from this central region, there are a number of hydrogen bonds (data not shown). There is a short stretch of antiparallel β -sheet (β 44–46 and α 77–75) which helps to hold the long loop of the β -subunit to the hairpin loops of the α -subunit. In addition, there are hydrophobic contacts between valine and leucine at β 44–45 and the triplet of phenylalanine residues (α 17, 18, 74) in the hairpin loop of the α -subunit. The complete association buries a total surface area of 4,525 Å² between the subunits: 2,134 Å² by α , 2,241 Å² by β , and 150 Å² by the carbohydrate on the α -subunit, specifically that of Asn 52.

An unexpected feature of the association is the role played by the loop from Cys 90 to Cys 110 in the β -subunit (Fig. 2*a*). In the heterodimer this loop is wrapped over the α -subunit while remaining covalently bonded to the β -subunit through the disulphide linkages (9–90, 26–110), and has the appearance of a seat belt (Fig. 3*a*). There is close association between the inner surface of the belt and the α -subunit with a short strand of parallel β -sheet between β (99–101) and α (53–57) forming part of the central β -barrel. The side chains of four α -subunit residues form a surface which interacts with either main chain or side chain atoms of the belt. There are hydrogen bonds between the side chains of α 35 Arg and β 106 His and the side chain of α 56 Glu and the main chain of β 104 Lys. Non-bonded interactions are

formed between α 37 Tyr, β 107 Pro and β 108 Leu and between α 54 Thr and β 99 Asp. There are no contacts specific to the hCG molecule that would prevent this seat-belt arrangement existing in all of the glycoprotein hormones.

Formation of Cys 93–100 is necessary for association with the α -subunit¹³ and a form of hCG β consisting of residues 8–100 associates only weakly with α ¹⁴. The disulphide Cys 26–110 is completed after the α/β association has occurred¹⁵, indicating that the seat-belt region is important in maintaining the integrity of the heterodimer.

Carbohydrate structure

In its native form, hCG is heavily glycosylated with complex carbohydrate making up ~34% by weight of the protein. *N*-linked carbohydrates occur at Asn 52 and Asn 78 on the α -subunit and at Asn 13 and Asn 30 on the β -subunit. In addition, the β -subunit has four *O*-linked carbohydrates at Ser 121, 127, 132 and 138. Treatment of the protein with anhydrous HF for one hour leaves the *O*-linked carbohydrate largely intact and truncates the asparagine-linked carbohydrate to predominantly Asn-(GlcNAc)₂ (refs 4, 16).

In the crystal structure, electron density is clear for two sugar residues on each of the α -subunit *N*-linked carbohydrates (Fig. 4*a*) and only one sugar residue is seen for those of the β -subunit. These fragments are sufficient to allow modelling of the complete carbohydrate (Fig. 4*b*). In the α -subunit, the carbohydrates are at the extremities of the molecule: Asn 52 is on the double-stranded loop and Asn 78 is near the end of the β -hairpins. The carbohydrate structure makes few hydrogen bonds to the protein. In the β -subunit, both *N*-linked carbohydrates are exposed on the outward faces of adjacent β -strands with Asn 13 and Asn 30 less than 7 Å apart (Fig. 3*a*). Some of the glycoprotein hormones have only one of these sites glycosylated. The only carbohydrate that interacts at the subunit interface is on α 52 Asn which contacts the β -subunit residues

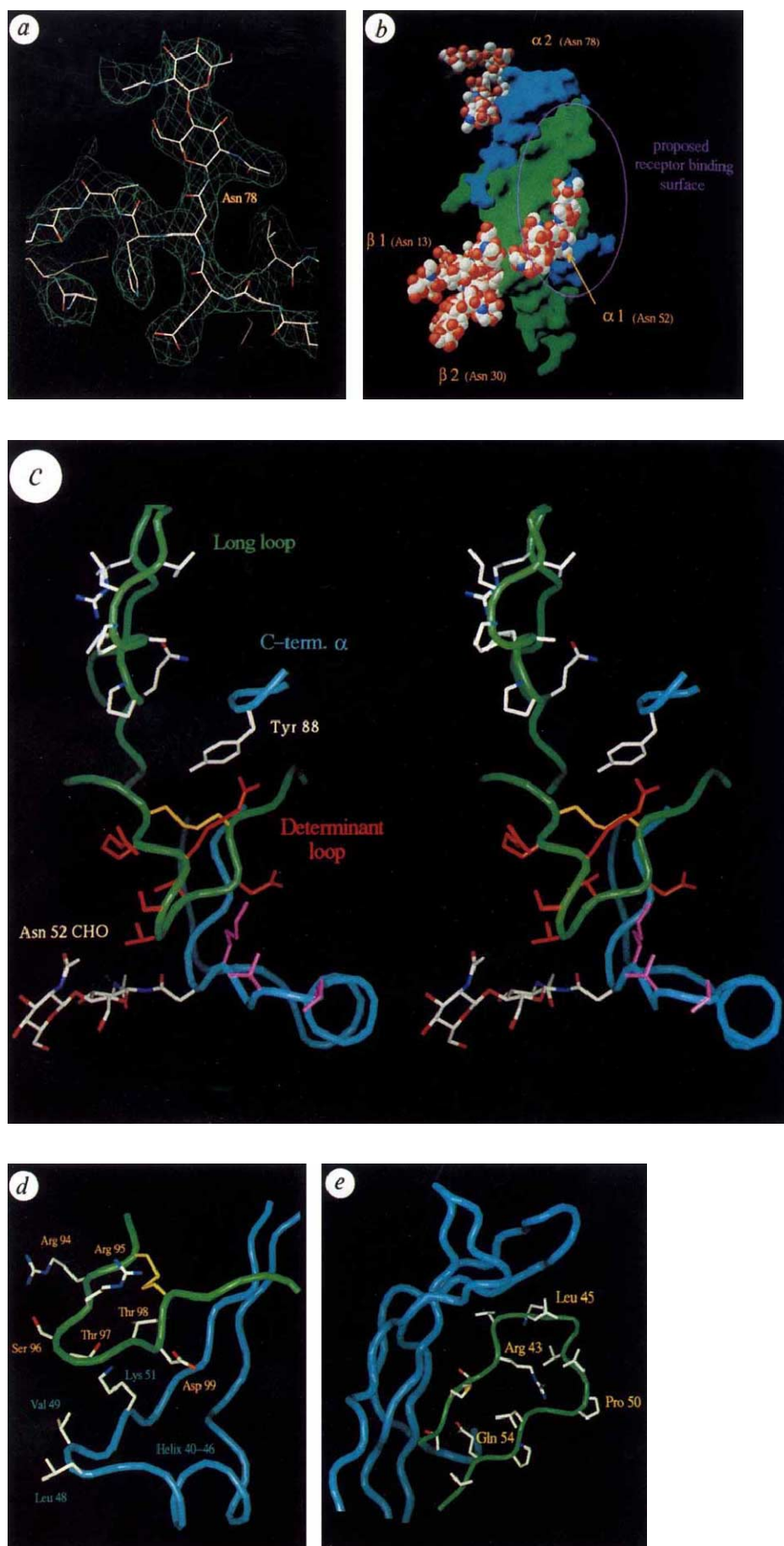


FIG. 4 *a*, Carbohydrate density at $\alpha 78$ Asn contoured at a 1.0σ level in the final $2|F_o - F_c|$ difference Fourier, showing clear density for two *N*-acetyl-glucosamine units. *b*, A model of the complete *N*-linked carbohydrate on the hCG molecule illustrates the extent of the carbohydrate structure. The protein is represented as surfaces (α , blue; β , green), with the carbohydrate represented as a space-filling model with atom colouring: N, blue; O, red; C, white. The carbohydrate was modelled by optimizing the fit of a complex biantennary carbohydrate from thrombin (coordinates provided by P. Martin) on the observed sugar residues at each of the glycosylation sites. The carbohydrate at $\alpha 78$ is far removed from the other three sites. The positions of the $\beta 13$ and $\beta 30$ carbohydrates are very close, forming a large bulk of carbohydrate. Some distal residues of the β carbohydrate could be close to one of the antennae of the $\alpha 52$ carbohydrate which is located at the dimer interface and is within the proposed receptor-binding domain. This sugar is best placed for biological function. *c*, A stereo view of the proposed receptor-binding site. The essential determinant loop ($\beta 93$ –100, residues in red) is centrally positioned with the β -subunit long loop (38–57, coloured by atom type) and Tyr 88 (white) of the α -subunit C-terminus above. Below is the carbohydrate (coloured by atom type) at $\alpha 52$ Asn, with the adjacent $\alpha 40$ –50 loop. Some invariant residues are coloured magenta. *d*, Side chains in the determinant loop $\beta 93$ –100, with colour coding for atom type. The adjacent helical region of the α -subunit shows the position of Leu 48, Val 49 and Lys 51. *e*, The unusual wedge-shaped $\beta 38$ –57 long loop extends from the hairpin loops of the α -subunit. Hydrophobic residues are predominantly surface-oriented; hydrophilic residues are either in the centre of the loop or buried in the α/β interface.

Tyr 59, Val 62, Phe 64 and Ala 83 as well as Thr 97, from the determinant loop.

Receptor binding and biological activity

A number of residues important for the activity of hCG have been identified through chemical modification, the use of synthetic peptides in competitive inhibition studies, and site-directed mutagenesis (reviewed in refs 2, 17, 18). As a result, several regions that contribute to receptor binding have been identified (Fig. 4).

The (β 93–100) determinant loop sequence (Fig. 4d). The importance of this sequence was recognized by Ward and Moore¹⁹, who proposed that the variable charge in this loop could act as a determinant of specificity, positive for LH/hCG and negative for FSH and TSH. The disulphide 93–100 which forms this loop is important for activity²⁰ and site-directed mutagenesis of either Cys 93 or Cys 100 in hCG β ²¹ yields mutant forms incapable of associating with the α -subunit.

The determinant loop is surface-oriented and held in place by a short stretch of parallel β -sheet between α 53–57 and β 99–101. The disulphide β 93–100 sits between the α 53–56 and β 50–57 strands, making non-bonded contacts with α 55 Ser and β 56 Val and constrains the loop to form a single turn of 3_{10} helix-like structure between β 94 Arg and β 98 Thr. This latter residue is buried at the interface with the α -subunit and its importance in subunit association has been demonstrated by mutagenesis²². Residues of the β -subunit implicated in receptor binding²² (Arg 94, Arg 95, Ser 96 and Asp 99) are accessible at the dimer surface. Thr 97, generally insensitive to mutagenesis, is in close contact with the carbohydrate of α 52 Asn.

The (β 38–57) sequence (Fig. 4e). Much attention has been focused on the longest inter-cysteine loop β 38–57. Synthetic peptides with this sequence stimulate steroidogenesis in rat Leydig cells²³ and strongly inhibit binding of whole hCG²³.

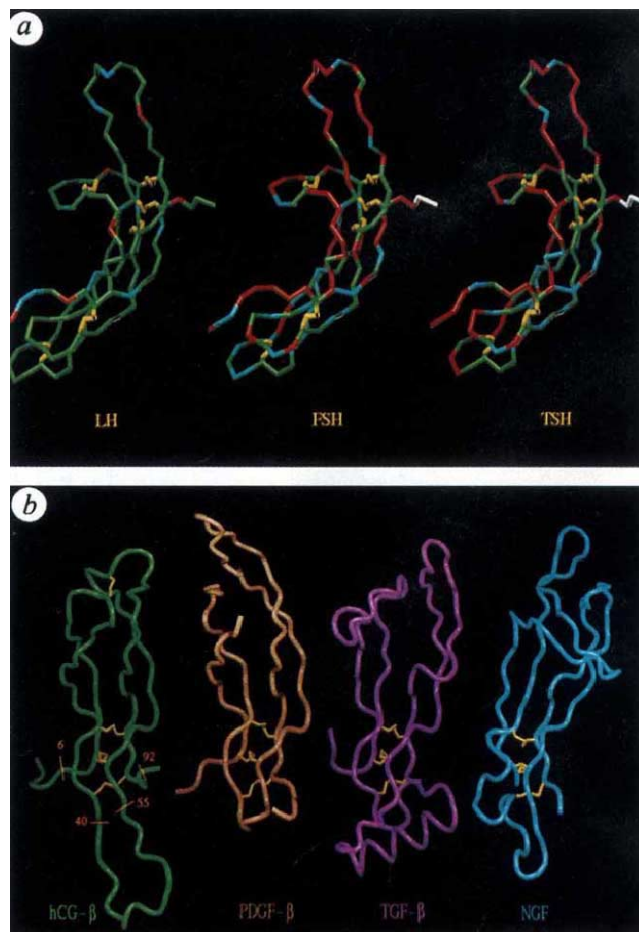
The crystal structure shows an unusual loop with residues 40–46 buried at the α/β interface (Fig. 3). Mutagenesis of Arg 43 to Leu (as in FSH) in hCG²⁴, or replacement of it by Ala or Asp in synthetic (38–57) peptides²⁵ either significantly diminishes binding or eliminates it. A natural mutation in LH of Gln 54 to Arg causes hypogonadism and the mutant hormone does not bind to its receptor²⁶. Residues 47–53 are predominantly hydrophobic and exposed at the surface, forming a wedge-shaped extrusion (Fig. 4e). This unusual structure, by virtue of its composition and proximity to the determinant loop, is likely to be important in receptor binding. The orientation of Arg 43 suggests that it may stabilize the conformation of the loop, although direct interaction with the receptor cannot be excluded.

The (α 88–92) C-terminus. Three of the five C-terminal residues (Tyr-Tyr-His-Lys-Ser-COOH), are identical in 13 of the 18 available amino-acid sequences. Carboxypeptidase digestion of both hCG α and LH α , although not detrimental to subunit assembly, essentially eliminates receptor binding^{27,28}. hCG α lacking residues 89–92 forms an active heterodimer but with reduction of both receptor binding and steroidogenesis²⁹.

Although residues 89–92 could not be positioned, the C-terminus of the α -subunit should be located close to both the β 38–57 sequence loop and the determinant loop of the β -subunit, forming a composite receptor-binding site (Fig. 4c). The side chain of α 88 Tyr packs between the disulphides β 93–100 and α 59–87. Mutagenesis of this residue to Phe causes reduced binding affinity but increases the steroidogenic response at saturating concentrations²⁹. As it is unlikely to be involved directly in receptor binding, this result may be due to subtle differences in conformation induced in the final four C-terminal residues.

Other potential α -subunit residues. The α -subunit residues 40–50 contain the only helical structure in the protein. The helix consists of highly conserved residues and is adjacent to the determinant loop (Fig. 4c, d), suggesting a receptor-binding role. This

FIG. 5 a, A α trace of hCG- β , coloured according to variation in the amino-acid sequence of the other human glycoprotein hormones. A log-odds matrix⁵⁰ was used to define non-conservative (red) and conservative (cyan) sequence changes. Invariant residues are coloured green, deletions white. The conserved cystine side chains are also shown. b, Schematic drawings of structures of hCG- β (green) compared to PDGF- β (tan), TGF- β (magenta) and NGF (blue) (coordinates from the Brookhaven Protein Data Bank⁵¹; not given for residues 27–37 in PDGF- β). The N termini and C termini of hCG- β and TGF- β have been truncated to simplify the diagram. The structures share the same cystine-knot motif, imposing strong conformational restraints on associated residues. There are significant changes in the size and shape of the hairpin loops, but the largest differences are seen in the long loop. Positions where hCG- β is cleaved to form the β -core fragment are shown in red and highlight the similarity between this fragment and PDGF- β .



is supported by recent work³⁰ implicating Lys 44 in receptor binding. The structure further suggests that the invariant Leu 48 and Val 49, which protrude into solvent at a turn, and Lys 51, which extends up towards the determinant loop, also play a part. **Carbohydrate.** As well as modifying the rate of clearance of the glycoprotein hormone from the body, some glycosylation is essential for full biological activity³¹. The LH/CG receptor contains, within the extracellular domain, a sequence similarity to soybean lectin³², which suggests that there is a binding site for carbohydrate. HF-hCG binds to receptors with slightly increased affinity but with loss or reduction of potency³³. Studies on the function of the carbohydrate show that the *N*-glycosylation sites are more important than the *O*-linked sites; also, glycosylation is not necessary for receptor recognition, and *N*-glycosylation of the α -subunit at Asn 52 is critical for signal transduction^{34,35}.

The carbohydrate at α 52 Asn is positioned at the dimer interface, about 4 Å away from the determinant loop (Fig. 4c). Its position, size and flexibility could be antagonistic to a peptide-based receptor binding, thereby explaining how removing carbohydrate increases receptor binding. Removal of both β -subunit *N*-glycosylation sites does not cause much loss of activity, but reduces the maximal steroidogenic activity³⁶. Although the two Asn residues are very close to each other and located 20 Å from α 52 Asn, modelling with the complete complex carbohydrate shows that some of the distal sugar residues on β 13, β 30 and α 52 could be close together (Fig. 4b).

Deglycosylated hCG can be converted back into an active form when bound to polyclonal antibodies^{31,37}, suggesting that a modified conformation is restored to the native form by the bound antibody³¹. As crystals of asialo-hCG are isomorphous with HF-hCG, there is no evidence in the crystal structure for conformational change caused by deglycosylation. The structure does support the alternative explanation³⁷ that the antibodies may mimic the steric interaction of the carbohydrate with the receptor.

The family of glycoprotein hormones

It is evident that the structures of the β -subunits of the other members of this family (LH, FSH, TSH) will be essentially simi-

lar to the structure described here. Figure 5a shows the positions of the non-conserved, conserved and invariant residues for each of the human hormones when compared to hCG. LH has only a few random non-conserved changes, as expected. FSH and TSH both have functionally distinct receptors and less sequence similarity. The non-conservative sequence changes are concentrated in the receptor-binding regions, namely the 38–57 long loop and the 93–100 determinant loop. There are unexpected changes in regions associated with the α -subunit, predominantly in the 100–110 seat belt and 39–47 region of the long loop. Conserved and invariant residues are concentrated in the two hairpin loops and the cystine knot, implying that these are important in maintaining the fold. Charged surface residues are also conserved in the hairpin loops, indicating that they may serve some common function for these hormones.

Cystine-knot growth factors

The correct assignment of the disulphide bridges and the determination of the tertiary fold of hCG reveals, surprisingly, that the glycoprotein hormones are members, with NGF, TGF- β and PDGF- β , of the superfamily of cystine-knot growth factors. In this family the proteins form either homodimers or heterodimers with homologous cystine-knot subunits. In contrast, the glycoprotein hormones form only heterodimers between a common α -subunit and different β -subunits. Structural comparisons with TGF- β , PDGF- β and NGF show that the cystine-knot motif allows variation in the size and conformation of the constituent loops, but there is a pronounced similarity between part of hCG- β and PDGF- β (Fig. 5b). Whereas hCG is found mainly as a heterodimer in the early stages of pregnancy, in the later stages both subunits are found in large quantities as separate molecules. Free α -subunit is more heavily glycosylated and has hormonal activity³⁸. Smaller forms of the β -subunit that have been considered degradation products are found. The smallest of these, β -core³⁹, is composed of two polypeptides, β 6–40 and β 55–92, which arise from excision of 15 residues from the long loop as well as truncation at the N and C termini. The structural similarity of this truncated β -molecule with PDGF is striking and suggests that the circulating β -core molecule has a defined tertiary structure and perhaps a biological function. □

Received 2 March; accepted 12 May 1994.

- Manjunath, P. & Sairam, M. R. *J. biol. Chem.* **257**, 7109–7115 (1982).
- Ryan, R. J. et al. *Rec. Prog. Horm. Res.* **43**, 383–429 (1987).
- Harris, D. C., Machin, K. J., Evin, G. M., Morgan, F. J. & Isaacs, N. W. *J. biol. Chem.* **264**, 6705–6706 (1989).
- Lustbader, J. W. et al. *Biochemistry* **28**, 9239–9243 (1989).
- Brunger, A. T. *XPLOR Manual* (Yale Univ. Press, New Haven, 1992).
- Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. *Proteins* **12**, 345–364 (1992).
- Mise, T. & Bahl, O. P. *J. biol. Chem.* **255**, 8516–8522 (1980).
- Mise, T. & Bahl, O. P. *J. biol. Chem.* **256**, 6587–6592 (1981).
- Murray-Rust, J. et al. *Structure* **1**, 153–159 (1993).
- McDonald, N. Q. et al. *Nature* **354**, 411–414 (1991).
- Schlunegger, M. P. & Grutter, M. G. *J. molec. Biol.* **231**, 445–458 (1993).
- Oefner, C., D'Arcy, A., Winkler, F. K., Eggiman, B. & Hosang, M. *EMBO J.* **11**, 3921–3926 (1992).
- Chen, F. & Puett, D. *J. biol. Chem.* **266**, 6904–6908 (1991).
- Puett, D., Huang, J. & Xia, H. in *Glycoprotein Hormones* (eds Lustbader, J. W., Puett, D. & Ruddon, R. W.) 122–134 (Springer, New York, 1994).
- Huth, J. R., Mountjoy, K., Perini, F. & Ruddon, R. W. *J. biol. Chem.* **267**, 8870–8879 (1992).
- Keutmann, H. T., McIlroy, P. J., Bergert, E. R. & Ryan, R. J. *Biochemistry* **22**, 3067–3072 (1983).
- Sairam, M. R. in *Hormonal Proteins and Peptides* (ed. Li, C.H.) **XI**, 1–79 (Academic, New York, 1983).
- Keutmann, H. T. *Molec. cell. Endocr.* **86**, c1–c6 (1992).
- Ward, D. N. & Moore, W. T. in *Animal Models for Research on Contraception and Fertility* (ed. Alexander, N. J.) 151–161 (Harper and Row, New York, 1979).
- Keutmann, H. T., Mason, K. A., Kitzmann, K. & Ryan, R. J. *Molec. Endocr.* **3**, 526–531 (1989).
- Suganuma, N., Matzuk, M. M. & Boime, I. *J. biol. Chem.* **264**, 19302–19307 (1989).
- Huang, J., Ujihara, M., Xia, H., Chen, F., Yoshida, H. & Puett, D. *Molec. cell. Endocr.* **90**, 211–218 (1993).
- Keutmann, H. T. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2038–2042 (1987).
- Chen, F. & Puett, D. *Biochemistry* **30**, 10171–10175 (1991).
- Keutmann, H. T. et al. *Biochemistry* **27**, 8939–8944 (1988).

- Weiss, J. et al. *New Engl. J. Med.* **326**, 179–183 (1992).
- Merz, W. E. *Eur. J. Biochem.* **101**, 541–553 (1979).
- Parsons, T. F. & Pierce, J. G. *J. biol. Chem.* **254**, 6010–6015 (1979).
- Chen, F., Wang, Y. & Puett, D. *Molec. Endocr.* **6**, 914–919 (1992).
- Xia, H., Chen, C. & Puett, D. *Endocrinology* **134**, 1768–1770 (1994).
- Sairam, M. R. *FASEB J.* **3**, 1915–1926 (1989).
- McFarland, K. C. et al. *Science* **245**, 494–499 (1989).
- Chen, H.-C., Shimohigashi, Y., Dufau, M. L. & Catt, K. J. *J. biol. Chem.* **257**, 14446–14452 (1982).
- Sairam, M. R. & Bhargavi, G. N. *Science* **229**, 65–67 (1985).
- Matzuk, M. M., Keene, J. L. & Boime, I. *J. biol. Chem.* **264**, 2409–2414 (1989).
- Sairam, M. R. & Jiang, L. G. *Molec. cell. Endocr.* **85**, 227–235 (1992).
- Rebois, R. V. & Liss, M. T. *J. biol. Chem.* **262**, 3891–3896 (1987).
- Blithe, D. L. in *Glycoprotein Hormones* (eds Lustbader, J. W., Puett, D. & Ruddon, R. W.) 156–166 (Springer, New York, 1994).
- Birken, S. et al. *Endocrinology* **123**, 572–583 (1988).
- CCP4 (SERC) Collaborative Computing Project 4, Daresbury Lab. Warrington, UK, 1992).
- Kabsch, W. *J. appl. Crystallogr.* **21**, 916–924 (1988).
- Sheldrick, G. M. *Acta crystallogr.* **46**, 467–473 (1990).
- Wang, B. C. *Meth. Enzym.* **115**, 90–112 (1985).
- Bricogne, G. & Gilmore, C. *Acta crystallogr.* **A46**, 284–297 (1990).
- Read, R. J., *Acta crystallogr.* **A42**, 140–149 (1986).
- Jones, T. A. *Meth. Enzym.* **115**, 157–171 (1985).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110–119 (1991).
- Holmgren, A. & Morgan, F. *Eur. J. Biochem.* **70**, 377–383 (1976).
- Nicholls, A., Sharp, K. A. & Honig, B. *Proteins* **11**, 281–296 (1991).
- Jones, D. T., Taylor, W. R. & Thornton, J. M. *Compl. appl. Biosci.* **8**, 275–282 (1992).
- Bernstein, F. C. et al. *J. molec. Biol.* **112**, 535–542 (1977).

ACKNOWLEDGEMENTS. We thank G. MacDermott and A. Freer for assistance with data collection; P. Emsley for help with computation; E. Dodson for advice on phase refinement; S. Birken for preparing some of the protein material; C. Carter and S. Xiang for advice on entropy calculations; and the staff at the SERC Daresbury Synchrotron Radiation Source for assistance. This work was supported by the MRC, in part by the WHO Task Force on Vaccines for Fertility Regulation and by a NATO Travel Award.

A possible fossil galaxy group

T. J. Ponman*, D. J. Allan*, L. R. Jones†, M. Merrifield†, I. M. McHardy†, H. J. Lehto‡ & G. A. Luppino§

* School of Physics and Space Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

† Physics Department, University of Southampton, Southampton SO9 5NH, UK

‡ Tuorla Observatory, Turku University Observatory, FIN-21500 Piikkiö, Finland

§ University of Hawaii, Institute for Astronomy, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

MOST galaxies, including our own, are located in groups¹. In the most compact galaxy groups, the members are separated on the sky by only a few galactic radii, and numerical simulations² suggest that such systems will merge to form a single elliptical galaxy (a 'fossil group') in a few billion years. Recent observations^{3,4} have shown that some compact groups are surrounded by X-ray-emitting haloes of hot gas, and have suggested that they contain substantial amounts of dark matter. An elliptical galaxy formed by the merger of such a group will retain its halo⁴, which is unaffected by merging. Using recent X-ray observations from Rosat we have searched for fossil groups, and we report here the discovery of a possible candidate at a redshift of 0.171. This candidate has a high X-ray luminosity, comparable to those of the brighter compact groups⁵, and appears similar to the giant elliptical galaxies at the centres of clusters, yet it is apparently isolated. Its optical properties are also consistent with an origin as the merged remains of a typical compact group.

The Rosat extended deep survey⁶ involved a sequence of overlapping pointed observations, spaced at 10' intervals, covering a 4° strip of sky near the North Galactic Pole. Over 210 X-ray sources were detected in this strip, most of which are expected to be distant quasars⁷. To pick out possible fossil groups, we searched each individual field for sources using a likelihood source searching program (D.J.A., T.J.P., R. D. Jeffries, manuscript in preparation), and then tested for significant source extension by calculating the likelihood for a series of increasingly blurred point spread functions (PSFs). We further reduced the list of candidates by discarding those with soft spectra, typical of quasars. Comparison with the Palomar sky survey plates produced a single candidate, RX J1340.6+4018, which coincided with an apparently isolated elliptical galaxy.

The X-ray image (Fig. 1) contains 430 detected photons, is elongated north-south and extends (even across its short dimension) significantly beyond the instrument PSF (Fig. 2a). An elliptical core-index model for the X-ray surface brightness $S(x, y) = S_0[1 + (x/a)^2 + (y/b)^2]^{-\alpha}$ (where x and y are the major and minor axes, a and b are the corresponding core radii, and α is a power law index), commonly used to fit the X-ray emission from galaxy clusters, gives a reasonable representation within 1.8' of the optical position (allowing for blurring by the PSF). The fit gives an X-ray centroid of RA(2000) 13 h 40 min 33.4 s, dec(2000) 40° 17' 48", an axis ratio of $2.7^{+0.6}_{-0.5}$, and position angle (east of north) $PA = 6^\circ \pm 4^\circ$ (all errors 1 σ). The core radius trades off against the index, but lies in the range $a = 0.6$ – $1.7'$ (major axis). Spectra derived from background-subtracted data within $r = 1.2'$ of the X-ray centroid (enclosing 60% of the total source flux) in the three contributing pointings, were fitted simultaneously with a hot plasma model⁸, giving temperature $T = 0.92 \pm 0.08$ keV and metallicity $Z = 0.36^{+0.15}_{-0.11}$ solar, with absorbing hydrogen column fixed at the value $n_H = 7.4 \times 10^{19} \text{ cm}^{-2}$ derived from radio surveys⁹.

R- and V-band CCD (charge-coupled device) images of the system (Fig. 1) were obtained with the Nordic optical telescope. The apparent V magnitude of the large elliptical is 17.2, and an

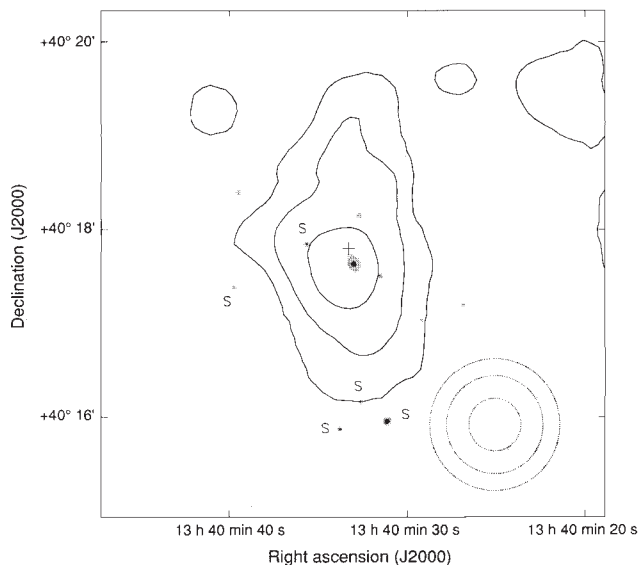


FIG. 1 X-ray contours superimposed on the (smaller) optical R-band image. All optical sources visible are galaxies, apart from five foreground stars marked 'S', the second brightest galaxy in the field being 2.3 mag fainter than RX J1340.6+4018. The 0.1–2.3 keV PSPC (Positron Sensitive Proportional Counter) image has been assembled as a mosaic from three pointings of ~6,500 s each, smoothed lightly to suppress noise, and contoured at logarithmic intervals of 2 in surface brightness. Background flux has been subtracted from each image as described in ref. 4. The X-ray centroid from fitting an elliptical model is marked with a cross, and coincides with the giant elliptical galaxy to within the Rosat attitude error of ~10". The instrument point spread function, smoothed and contoured in the same way as the data, is shown dotted (lower right) for comparison.

analysis of its optical surface brightness using elliptical isophote fitting¹⁰ shows that it is well represented by a de Vaucouleurs profile (Fig. 2b) with an effective radius $r_e = 5.6''$. No significant deviations are apparent from an elliptical model with ellipticity $e = 0.25$ (axis ratio 1.33) and $PA = 30^\circ$. The only unusual feature apparent is a tendency for the galaxy to redden with increasing radius—the opposite of the normal colour trend in elliptical galaxies¹¹. This effect ('red upturns') has been observed¹² previously in a number of central dominant galaxies in galaxy clusters, though its origin is unclear.

An optical spectrum obtained with the University of Hawaii 88-in telescope (Fig. 3), established the redshift of RX J1340.6+4018 as $z = 0.171$. Using $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $q_0 = 0.5$, the absolute V magnitude corrected for cosmological effects (K-correction) is $M_V = -23.5$, and its effective radius is 21 kpc. Such large size and luminosity are rare in field galaxies¹³, but are typical of the giant ellipticals (gE or D galaxies) which often dominate the centres of galaxy clusters^{14,15}. Deeper imaging is required to establish whether RX J1340.6+4018 has the extended low surface brightness envelope characteristic of centrally dominant (cD) galaxies.

The bolometric X-ray luminosity of RX J1340.6+4018, based on the redshift and the fitted X-ray spectrum, is $L_X = 4.5 \times 10^{43} \text{ erg s}^{-1}$. The X-ray core radius of 0.6–1.7', corresponds to 140–390 kpc, and emission is detected (Fig. 2a) out to a radius of ~500 kpc.

A numerical model of the emitting plasma cloud was constructed, in which electron density was parameterized as $n_e(r) = n_e(0)[1 + (r/r_c)^2]^{-\gamma}$, with temperature and metallicity fixed at the values determined from the integrated spectrum. This was fitted simultaneously to the spectral images within 1.5' of the X-ray centroid from each of the three pointings, allowing for the spatial and spectral response functions of the instrument, as described in ref. 4. The implicit assumption of spherical symmetry is

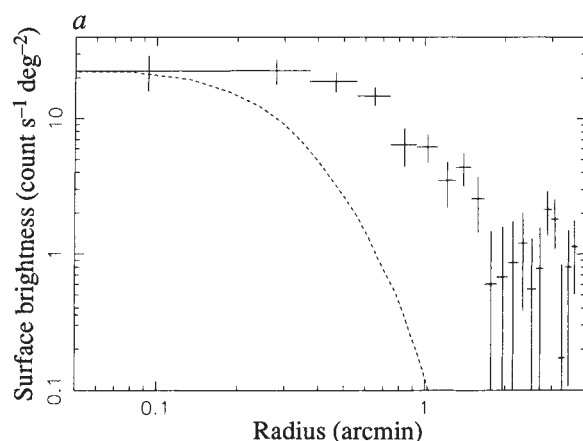
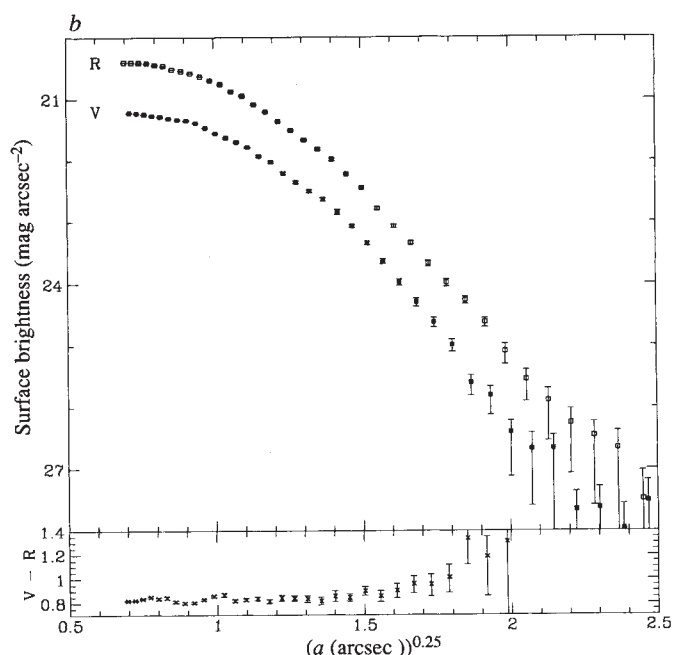


FIG. 2 a, Radial profile (data points with error bars) of the unsmoothed X-ray image about its centroid, with the appropriate PSPC point spread function (given the observed source spectrum and the off-axis angles of the three pointings) overlaid (dotted line). b, Profiles of optical surface brightness (top) and colour ($V-R$, bottom) along the major axis, produced by fitting elliptical isophotes¹⁰ to the V- and R-band images. The profiles are well represented by de Vaucouleurs models ($\log S \propto r^{0.25}$) outside the region affected by the 1.6'' seeing.



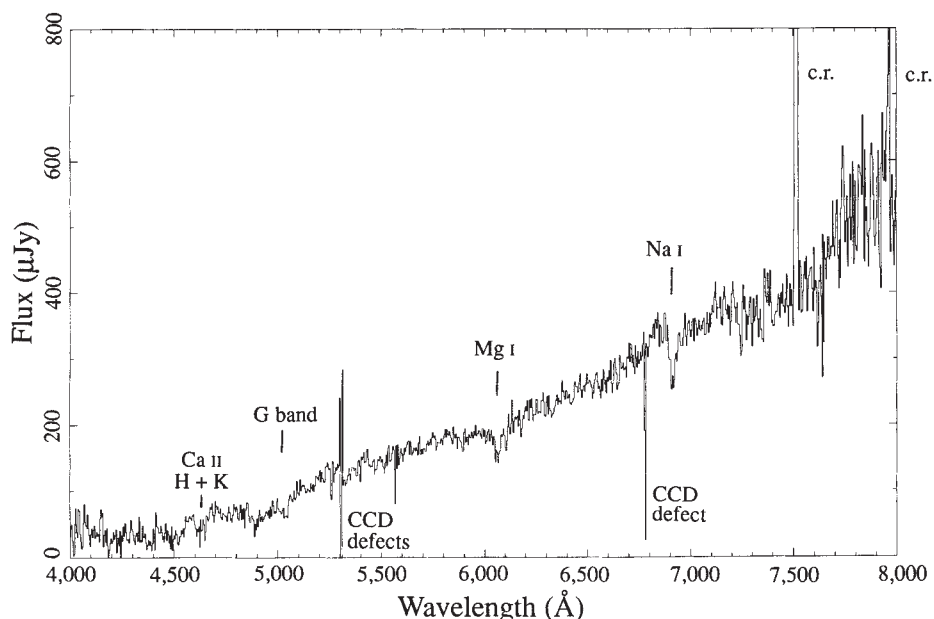
unfortunate, but has little impact on the main results (for example the density parameters agree well with the elliptical surface brightness fits). The best fitting model has $n_c(0) = 2.5 \times 10^{-3} \text{ cm}^{-3}$, $r_c = 0.8'$ and $\gamma = 1.5$, which implies a central cooling time of 9×10^9 yr.

The optical morphology of the elliptical galaxy strongly suggests that the system is relaxed, so that the gas should be in hydrostatic equilibrium and its pressure profile may be used to trace the gravitational potential. The gravitating mass within $r = 1.5' \equiv 340 \text{ kpc}$ is $M_g = 2.8 \times 10^{13} M_\odot$ (where $M_\odot$ is the solar mass), and the gas constitutes 9% of this. Given the absolute magnitude derived above, the mass/light (M/L) ratio of the system, within $r = 1.5'$, is ~ 130 solar units. The main error in the mass estimate arises from the fact that the density index (as with the surface brightness fits) trades off against core radius. The resulting allowed band in M/L is 90–260 solar units. The assumption that the gas is isothermal is unlikely to introduce an

error as large as this. If the nearest galaxies on the sky to RX J1340.6+4018 are bound to the system (their redshifts are unknown), this would increase L by $\sim 20\%$, whilst extrapolation of the mass distribution beyond our statistical cutoff at $r = 1.5'$ could increase M considerably. We conclude that $M/L \gtrsim 100$, so the gas is bound to the system by dark matter.

A deep radio map obtained with the Very Large Array (VLA) showed an unresolved ($r < 7''$) source with a flux of $2.5 \pm 0.5 \text{ mJy}$ at 1.5 GHz, coincident with the galaxy. This flux corresponds to a power of $3.5 \times 10^{30} \text{ erg s}^{-1} \text{ Hz}^{-1}$ at 1.5 GHz, which is considerably brighter than the normal elliptical galaxies surveyed by Fabbiano *et al.*¹⁶, though it is also more optically luminous. It is comparable to typical 'radio loud' cD galaxies found in cluster cores¹⁷. Most, but not all, of these are associated with cooling flows¹⁷. The central cooling time derived above suggests that RX J1340.6+4018 could contain an unresolved cooling flow, however the small size ($r \lesssim 60 \text{ kpc}$) and luminosity

FIG. 3 Low-resolution (12-Å) spectrum of the core of the galaxy, corrected for atmospheric absorption. The G band, Mg I and Na I absorption features seen in the spectra of G stars, is common in spectra of elliptical galaxies, give a redshift of $z = 0.171 \pm 0.001$. Cosmic ray (c.r.) and other spurious features are marked.



($L < 10^{43} \text{ erg s}^{-1}$) allowed for this indicate that it cannot be depositing gas at more than $\sim 30 M_{\odot} \text{ yr}^{-1}$. The optical spectrum shows no sign of the emission lines which would indicate an active nucleus, nor of the H α emission which often accompanies substantial cooling flows¹⁸ (our 3σ upper limit is $L(\text{H}\alpha) < 2 \times 10^{40} \text{ erg s}^{-1} \text{ arcsec}^{-2}$ in the galaxy core).

It is apparent from its X-ray properties that RX J1340.6+4018 is not a normal elliptical galaxy. Both the luminosity and extent of the hot gas are far larger than the hot gaseous haloes seen in field elliptical galaxies, which have¹⁹ $L_X < 10^{42} h_{50}^{-2} \text{ erg s}^{-1}$ (where $h_{50} \equiv H_0/50 \text{ km s}^{-1} \text{ Mpc}^{-1}$), radii $r \lesssim 100 h_{50}^{-1} \text{ kpc}$, and involve a gas mass $M_{\text{gas}} \lesssim 10^{11} M_{\odot}$.

Despite its X-ray luminosity and its similarity to a cluster-dominant galaxy, it is clear that RX J1340.6+4018 does not reside near the centre of a cluster. There is a total of 24 galaxies in the field of Fig. 1 (which has a diameter of 850 kpc at $z = 0.171$) with $R < 21$, compared with 20 predicted²⁰ from the sky-averaged density, and there is no enhancement centred on the giant elliptical. Also, the temperature of the X-ray emitting gas is lower than that expected in even very poor clusters^{21,22}.

Recent observations⁵ with Rosat have shown that gas at $\sim 10^7 \text{ K}$ is a common feature of compact galaxy groups, though its X-ray luminosity ranges over more than two orders of magnitude, $L_X \approx 10^{41} - (3 \times 10^{43}) h_{50}^{-2} \text{ erg s}^{-1}$. The X-ray properties of RX J1340.6+4018 are very similar to the brightest compact groups. For example, the hot gas in HCG62, has⁴ $L_X = 1.1 \times 10^{43} h_{50}^{-2} \text{ erg s}^{-1}$, $r \gtrsim 500 h_{50}^{-1} \text{ kpc}$ and $M_{\text{gas}} \gtrsim 2.5 \times 10^{12} M_{\odot}$. The dynamical mass of HCG62 ($2.7 \times 10^{13} M_{\odot}$ within $r = 500 \text{ kpc}$) is also similar to that inferred in RX J1340.6+4018.

It appears, then, that RX J1340.6+4018 is a giant elliptical galaxy which has been formed in an environment similar to that of an X-ray-bright galaxy group. The possibility that the galaxy has formed directly from the cooling gas seems unlikely for several reasons. The current cooling rate ($\lesssim 30 M_{\odot} \text{ yr}^{-1}$) is probably insufficient to accumulate a galaxy with an optical luminosity of $2.1 \times 10^{11} L_{\odot}$ within the age of the Universe. Moreover, studies of the properties of central galaxies embedded in the much stronger cooling flows seen in many clusters have shown that the cooling gas cannot be converted into stars with anything like a normal initial mass function^{23,24}.

The most natural explanation for RX J1340.6+4018 is that it is the merged remains of the galaxies which previously constituted the group. Its X-ray properties (luminosity, temperature, metallicity and size) are similar to those of the brighter compact groups⁵ and its optical luminosity is near the median for such groups²⁵. The similarity of RX J1340.6+4018 to cluster-dominant galaxies therefore lends support to the idea^{23,26} that these giant galaxies originate from merging within subclusters, which later combine to form rich clusters.

The X-ray ellipticity of 0.6 is much larger than is typical for the X-ray emission from normal ellipticals, and is considerably greater than that of the optical galaxy—implying that the extended distribution of dark matter that defines the gravitational potential is more elongated than the distribution of stars. N -body simulations have shown²⁷ that collapse of a prolate configuration (such as a filament) leads to a virialized distribution with an anisotropic velocity dispersion, the eccentricity of which increases with radius. The misalignment of 24° between the optical and X-ray position angles is larger than that typically seen in galaxy clusters²⁸, and may reflect the action of dissipative processes which have affected the galaxies but not the dark matter. Higher spatial resolution X-ray observations with the Rosat High Resolution Imager will allow a more detailed study of the X-ray morphology, and also a better determination of the gas cooling rate.

The fact that RX J1340.6+4018 was identified in a survey covering (to the depth that our technique would have been successful) only $\sim 10^6 \text{ Mpc}^3$, suggests that many more such systems will become available for study in the near future. These systems may be found both through Rosat studies, and from the discov-

ery, in deep-galaxy redshift surveys, of highly luminous elliptical galaxies in moderately low-density environments. $\square$

Received 20 December 1993; accepted 11 April 1994.

1. Tully, R. B. *Astrophys. J.* **321**, 280–304 (1987).
2. Barnes, J. E. *Nature* **338**, 123–126 (1989).
3. Mulchaey, J. S., Davis, D. S., Mushotzky, R. F. & Burstein, D. *Astrophys. J.* **404**, L9–L12 (1993).
4. Ponman, T. J. & Bertram, D. *Nature* **363**, 51–53 (1993).
5. Ponman, T. in *Cosmological Aspects of X-ray Clusters of Galaxies* (ed. Seitter, W.) (NATO ASI, Kluwer, Dordrecht, in the press).
6. Jones, L. R. et al. in *Observational Cosmology* (eds Chincarini, G. et al.) 468–469 (ASP, San Francisco, 1993).
7. Shanks, T. et al. *Nature* **353**, 315–320 (1991).
8. Raymond, J. C. & Smith, B. W. *Astrophys. J. Suppl. Ser.* **35**, 419–439 (1977).
9. Stark, A. A. et al. *Astrophys. J. Suppl. Ser.* **79**, 77–104 (1992).
10. Jedrzejewski, R. I. *Mon. Not. R. astr. Soc.* **226**, 747–768 (1987).
11. Peletier, R. F. et al. *Astr. J.* **100**, 1091–1142 (1990).
12. McNamara, B. R. & O'Connell, R. W. *Astrophys. J.* **393**, 579–610 (1992).
13. Efsthathiou, G., Ellis, R. S. & Peterson, B. A. *Mon. Not. R. astr. Soc.* **232**, 431–461 (1988).
14. Hoessel, J. G. & Schneider, D. P. *Astr. J.* **90**, 1648–1664 (1985).
15. Schombert, J. M. *Astrophys. J. Suppl. Ser.* **60**, 603–693 (1986).
16. Fabbiano, G. et al. *Astrophys. J.* **312**, 111–121 (1987).
17. Burns, J. O. *Astr. J.* **99**, 14–30 (1990).
18. Heckman, T. M. et al. *Astrophys. J.* **338**, 48–77 (1989).
19. Fabbiano, G. A. *Rev. Astr. Astrophys.* **27**, 87–138 (1989).
20. Jones, L. R., Fong, R., Shanks, T., Ellis, R. S. & Peterson, B. A. *Mon. Not. R. astr. Soc.* **249**, 481–497 (1991).
21. Kriss, G. A., Cioffi, D. F. & Canizares, C. R. *Astrophys. J.* **272**, 439–448 (1983).
22. David, L. P. et al. *Astrophys. J.* **412**, 479–488 (1993).
23. Edge, A. C. *Mon. Not. R. astr. Soc.* **250**, 103–110 (1991).
24. Fabian, A. C. *Rev. Astr. Astrophys.* (in the press).
25. Hickson, P., de Oliveira, C. M., Huchra, J. P. & Palumbo, G. G. C. *Astrophys. J.* **399**, 353–367 (1992).
26. Schombert, J. M. *Astrophys. J. Suppl. Ser.* **64**, 643–666 (1987).
27. Rhee, G. & Roos, N. *Mon. Not. R. astr. Soc.* **243**, 629–636 (1990).
28. Carter, D. & Metcalfe, N. *Mon. Not. R. astr. Soc.* **191**, 325–337 (1980).

ACKNOWLEDGEMENTS. The ROSAT Extended Deep Survey is a collaborative project involving the Universities of Southampton, Birmingham, London (Mullard Space Science Laboratory and Imperial College) and Oxford. We thank A. P. Willmore, S. Zepf, C. Crawford and N. Metcalfe for discussions. Data reduction and analysis were performed on the Starlink nodes at Birmingham and Southampton.

Scattering and absorption of surface electron waves in quantum corrals

E. J. Heller*, M. F. Crommie†‡, C. P. Lutz† & D. M. Eigler†

* Department of Physics and Harvard-Smithsonian Observatory, 17 Oxford Street, Harvard University, Cambridge, Massachusetts 02138, USA

† IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, California 95120–6099, USA

STANDING-WAVE patterns in electron density have been seen recently^{1–4} in images of the surfaces of noble metals obtained with the scanning tunnelling microscope (STM). These patterns are due to the scattering of surface electrons off impurities and step edges. By assembling specific enclosed structures of adatoms ('quantum corrals') using the STM, one can generate standing waves of particular geometries³. Here we describe a theory of the scattering process, which allows us to predict the standing-wave patterns of an arbitrary corral geometry with great accuracy. We can use the theory to examine the scattering properties of the atoms in the corral walls. We find that iron atoms assembled on the (111) surface of copper act as 'black dots', soaking up all of the electron wave amplitude impinging on them. A scattered wave is generated nonetheless, but this behaviour means that the corral walls are only 25% reflective. In an acoustic analogy, the corral is therefore a rather quiet chamber.

‡ Present address: Department of Physics, Boston University, Boston, Massachusetts 02215, USA.

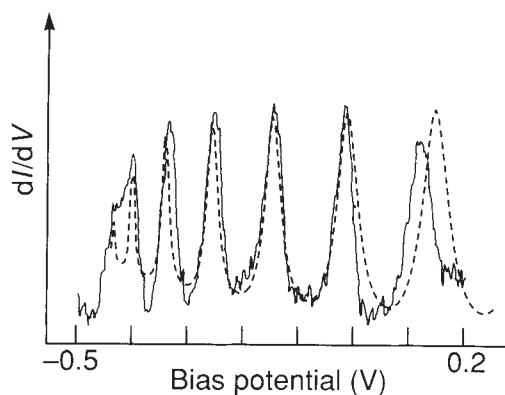


FIG. 1 The experimental (solid line) and theoretical (dashed line) voltage dependence of dI/dV , with the top of the STM located at the centre of a 88.7-Å diameter, 60-atom circle of Fe atoms on a copper surface (see text). The experiment has had a smooth background removed.

Crommie *et al.*³ explained the electron density distribution inside quantum corrals using a particle-in-a-box model. This model, although simple and qualitatively correct, could not account for the density distribution outside the corral, for the widths of the resonances seen in spectroscopic measurements (the q -factors), or for arbitrary arrangements of atoms. No realistic account of the interaction between electrons and iron atoms was given. In the theory presented here, scattering by an adatom is parameterized by a complex scattering phase shift which, once determined by a fit to data, may be used quantitatively to predict the density distribution and energy spectrum (including linewidths) of an arbitrary arranged array of adatoms.

For positive bias voltages (similar arguments can be made for negative bias), electrons tunnel to the Cu(111) surface from the STM tip. This produces a region of enhanced electron amplitude which travels away from the tip along the surface. If the electron wave encounters a scattering centre, such as a step edge or adatom, it may be scattered and return to the region of the tip, where it will interfere constructively or destructively with the amplitude leaving the tip. The flux of electrons leaving the tip is proportional to the square of the wavefunction amplitude. The interference therefore will cause oscillations of the current as a function of either energy (determined by the bias voltage)

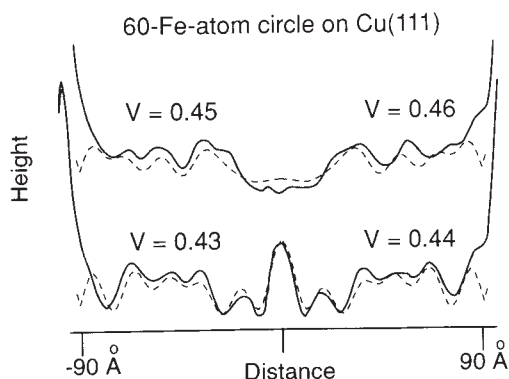


FIG. 2 The experimental data (solid lines) and the theoretical curves (dashed) for the tip height as a function of position across the diameter of the circle of atoms (Fig. 1), for low bias voltages.

or distance to the defect both of which affect the phase of the returning wave. The electron may scatter from several defects before returning, complicating the calculation of the interference but not the basic principle. If this simple picture is adopted, the problem is reduced to quantum scattering in two dimensions. We must find the outgoing and returning wave amplitude, and examine the interference of the two. The Fe adatoms are small compared to the wavelength of the electrons, which at the Fermi energy is about 50 atomic units (a.u., 1 a.u. = 0.529 Å). This means it is reasonable to assume only s -wave (spherically symmetric) scattering. The atoms are also remote from one another in the sense that the smallest distance between them is 9.5 Å, much bigger than the atomic diameter, allowing a multiple scattering expansion^{5,6}.

The electron amplitude arriving at the site of an adatom from the tip can be written as

$$a_T(r) = \left[\sqrt{\frac{2}{\pi k r}} e^{ikr - i\pi/4} \right] \quad (1)$$

where k is the wavevector of the electron waves, and r is the distance from the tip. This causes a scattered wave from the adatom

$$a_T(r_j)a(r) = a_T(r_j) \sqrt{\frac{2}{\pi k}} e^{i\pi/4} \frac{(e^{2i\eta_0} - 1)}{2i} \frac{e^{ikr}}{\sqrt{r}} \quad (2)$$

where η_0 is the phase shift and r_j is the distance to the adatom. A fraction of the incident electron wave may be absorbed at the adatom due to inelastic scattering and elastic scattering into bulk states. We shall call this the absorptive channel. To account for this absorption we use the standard procedure of taking η_0 to be complex, that is $\exp(2i\eta_0) \equiv \alpha_0 \exp(2i\delta_0)$, where α_0 and δ_0 are real. It is interesting that maximal 'black dot' attenuation, $\text{Im}(\eta_0) = \infty$ (where Im denotes the imaginary part), leads to $\alpha_0 = 0$ and the scattered wave

$$a(r) = \sqrt{\frac{2}{\pi k}} e^{i\pi/4} \frac{(\alpha_0 e^{2i\delta_0} - 1)}{2i} \frac{e^{ikr}}{\sqrt{r}} \rightarrow \sqrt{\frac{2}{\pi k}} e^{i\pi/4} \frac{e^{i\pi/2}}{2} \frac{e^{ikr}}{\sqrt{r}} \quad (3)$$

Thus there is still a scattered wave even for black-dot attenuation (a well known fact in scattering theory), and interference with the incident amplitude still happens.

The amplitude arriving at atom j' , having come from the tip and scattered off atom j , is then $a_T(r_j)a(r_{jj'})$, where $r_{jj'}$ is the distance between atoms j and j' . The total amplitude for scattering once and returning to the tip at $\mathbf{r}$ is $g_1(\mathbf{r}) = \sum_j a_T(r_j)a(r_j)$. The amplitude having scattered once from atom j and thence from j' before returning to the tip is $g_2(\mathbf{r}) = \sum_j \sum_{j'} a_T(r_j)a(r_{jj'})a(r_j')$. The outgoing tip amplitude is the Bessel function $J_0(kr)$, which at $r=0$ is equal to one. Thus the interference term between all the scattered amplitude and the tip amplitude is

$$\begin{aligned} \mathcal{I}(\mathbf{r}, k) &= 2 \text{Re} \left[\sum_{n=1}^{\infty} g_n(\mathbf{r}) \right] \\ &= 2 \text{Re} [\mathbf{a}_T(1 + \mathbf{A} + \mathbf{A}^2 + \dots)\mathbf{a}] \\ &= 2 \text{Re} [\mathbf{a}_T[1 - \mathbf{A}]^{-1}\mathbf{a}] \end{aligned} \quad (4)$$

where $\mathbf{a}_T$ and $\mathbf{a}$ are vectors whose length is N , the number of Fe adatoms, $\dots$, Re denotes the real part of the expression, and $\mathbf{A}$ is an $N \times N$ matrix. At each new position the distances from tip to the adatoms changes, so the vectors $\mathbf{a}_T$ and $\mathbf{a}$ change, but the matrix of atom-to-atom amplitudes $\mathbf{A}$ is fixed for constant bias voltage.

Most discussions of STM current are couched in terms of the local density of electron states (LDOS), $\text{LDOS}(\mathbf{r}, \varepsilon) = \sum_v |\psi_v(\mathbf{r})|^2 \delta(E_v - \varepsilon)$ where ε is the energy determined by the bias voltage and the Fermi level, and v labels the scattering

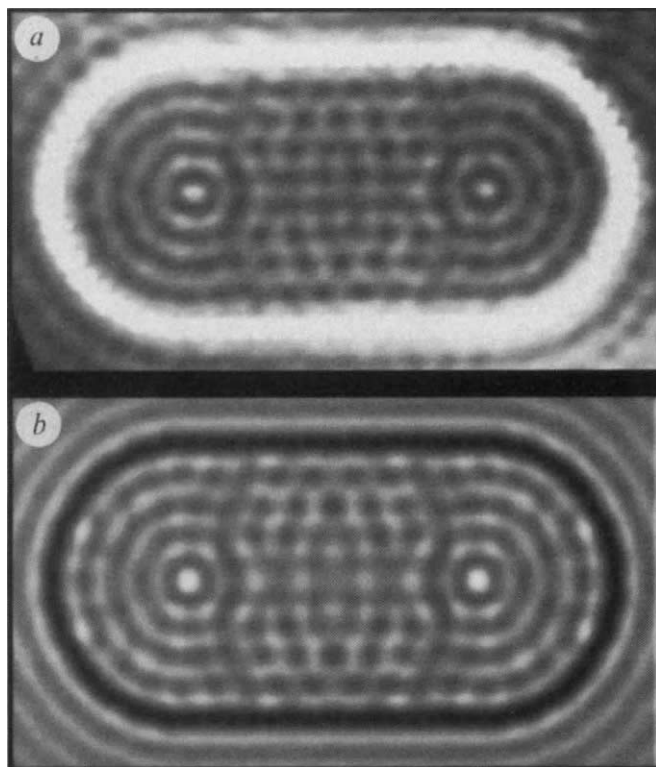


FIG. 3 Local density of electron states (LDOS) near E_F for a 76 Fe atom 'stadium' of dimensions $141 \times 285 \text{ \AA}$. *a*, Experiment, bias voltage 0.01 V ($\varepsilon = 0.45$); *b*, theory ($\varepsilon = 0.46$). The density near the centre of the Fe adatoms is not accounted for in the theory and appears black.

eigenstates including the presence of the adatoms. It may be shown quite easily (our unpublished results) that $\mathcal{J}(\mathbf{r}, k) \propto \text{LDOS}(\mathbf{r}, \varepsilon)$, where ε and k are related by the dispersion curve. This shows that our scattering approach is merely a way of getting the LDOS, and is completely consistent with the standard picture of STM tunnelling current proposed by Tersoff and Hamann⁷. The multiple-scattering approach is physically motivated as well as computationally suited to the presence of small, arbitrarily placed impurities in a two-dimensional system.

Equation (4) was used in the comparisons with experiment which follow. The adatom structures were assembled and studied using a low-temperature (4 K) ultra-high-vacuum STM, as described in refs 1 and 3. All STM topographs are constant-current images. For low bias voltages and small vertical tip excursions, features in the topographs are proportional to variations in the surface LDOS near E_F , the Fermi energy (ref. 6). For larger bias voltages measurement of dI/dV versus V (where I is the current and V is the bias voltage) gives the energy dependence of the LDOS⁸.

Consider a single isolated adatom. Here equation (4) reduces to $a_{\mathbf{r}}(r_j)a(r)$, yielding

$$\text{LDOS}(\mathbf{r}, \varepsilon) \propto \text{Re} \left[\frac{(\alpha_0 e^{2i\delta_0} - 1)}{2i} \frac{e^{2ikr}}{r} \right] \quad (5)$$

For no absorption ($\alpha_0 = 1$) we have $\text{LDOS}(\mathbf{r}, \varepsilon) \propto$

$\sin(\delta_0) \cos(2kr + \delta_0)/r$, which agrees with the single-atom expression given previously¹. In the previous work the observed pattern for scattering off a single Fe atom gave a phase shift near -80° just above E_F . But with no absorption the expression is invariant to a phase advance of π , so the phase could equally well have been near $+100^\circ$. The earlier work did not take into account the possibility of scattering into absorptive channels. The maximum 'black dot' attenuation gives an effective phase shift of $+90^\circ$ (due to the $\exp(i\pi/2)$ term in equation (3)) and also agrees with the single-atom data. The single-atom experiment thus can not distinguish between elastic and absorptive scattering. However the absorptive attenuation makes a crucial difference for many-atom experiments, which can be used to identify the correct scattering law. The black dot choice leaves no other parameter to specify except the wavevector k and this is inferred from the energy via the dispersion relation.

Figure 1 compares a measurement of dI/dV for the STM tip held above the centre of an $88.7\text{-\AA}$ radius circular corral to the theoretical LDOS using ($\alpha_0 = 0$). Note that except for the last peak, above the Fermi energy, the agreement is excellent. Figure 2 shows the experimental data (solid lines) and theoretical curves for tip height variations as a function of position across the diameter of the same corral, for various bias voltages. The theory gives excellent agreement with the experiment, provided that the wavevector k is adjusted to correspond to a small (0.01-eV) shift in the energy compared to experiment. This shift falls well within the limits of uncertainty on k set by the measured values of the surface-state band edge and effective mass¹. We use this same shift in all our calculations.

In Fig. 3 we show experimental and theoretical results for a 76-atom 'stadium' of dimension $141 \times 285 \text{ \AA}$. It is important, for this detailed agreement, to use the precise positions of the Fe atoms (which must register to the triangular grid of allowed binding sites on the Cu(111) face)³.

In the limit $\alpha_0 \rightarrow 0$, the elastic scattered amplitude has a 'cross-length' of $1/k$, and the cross length for the absorption of electrons out of the s -wave surface state is also $1/k$, as in the following equation.

$$\sigma = \frac{|\alpha_0 e^{2i\delta} - 1|^2}{k} + \frac{(1 - \alpha_0^2)}{k} \equiv \sigma_{\text{elastic}} + \sigma_{\text{absorption}} \quad (6)$$

These lengths each correspond to $\sim 4.5 \text{ \AA}$ at E_F , yielding a total cross length that is comparable to the average spacing between the atoms for the circle and the stadium, which is $\sim 9.3 \text{ \AA}$. We can estimate the fate of a plane wave normally incident on an infinite line of Fe atoms spaced (as in these experiments) by $\sim 10 \text{ \AA}$ and given the s -wave attenuation ($\alpha_0 = 0$). We find that $\sim 25\%$ of the incident amplitude is reflected and 25% transmitted, leaving 50% absorbed (presumably into the bulk states).

The probable explanation for most of the black-dot behaviour is scattering of surface state electrons into bulk Cu states. Although surface states of a perfect Cu crystal are orthogonal to bulk states, the existence of defects (such as the Fe adatoms) can provide coupling to the bulk⁹. The black-dot property will limit the collective behaviour of artificial nanostructures built from iron atoms; it is likely that other atoms will also act as black (or at least as 'grey') dots. By using thin-film substrates, however (thus removing the bulk states altogether), it may be possible to make the walls 'white' (that is, highly reflective). One could then explore the interesting realm of chaotic billiards on an atomic scale, or make artificial semiconductors with 'engineered' bandgaps. $\square$

Received 11 March; accepted 5 May 1994.

1. Crommie, M. F., Lutz, C. P. & Eigler, D. M. *Nature* **363**, 524–527 (1993).
2. Hasegawa, Y. & Avouris, P. *Phys. Rev. Lett.* **71**, 1071–1074 (1993).
3. Crommie, M. F., Lutz, C. P. & Eigler, D. M. *Science* **262**, 218–220 (1993).
4. Avouris, P., Lyo, I.-W., Walkup, R. E. & Hasegawa, Y. *J. Vac. Sci. Technol. B* (submitted).
5. Rodberg, L. S. & Thaler, R. M. *Introduction to the Quantum Theory of Scattering* (Academic, New York, 1967).
6. Gaspard, P. & Rice, S. A. *J. chem. Phys.* **90**, 2225–2262 (1989).

7. Tersoff, J. & Hamann, D. R. *Phys. Rev.* **B31**, 805–813 (1984).
8. Lang, N. D. *Phys. Rev.* **B34**, 5947–5950 (1986).
9. Kevan, S. D. *Phys. Rev. Lett.* **50**, 526–529 (1983).

ACKNOWLEDGEMENTS. E.J.H. thanks T. Kaxiras for discussion. We thank D. P. Brodbeck for help with the STM image rendering, and M. S. Sherwin for discussions. E.J.H. thanks the US National Science Foundation and the US Department of Energy for support of this work.

Conducting tin halides with a layered organic-based perovskite structure

D.B. Mitzi*, C. A. Feild*, W. T. A. Harrison† & A. M. Guloy†

* IBM T. J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA

† Department of Chemistry and Texas Center for Superconductivity, University of Houston, Houston, Texas 77204-5641, USA

THE discovery¹ of high-temperature superconductivity in layered copper oxide perovskites has generated considerable fundamental and technological interest in this class of materials. Only a few other examples of conducting layered perovskites are known; these are also oxides such as $(\text{La}_{1-x}\text{Sr}_x)_{n+1}\text{Mn}_n\text{O}_{3n+1}$ (ref. 2), $\text{La}_{n+1}\text{Ni}_n\text{O}_{3n+1}$ (ref. 3) and $\text{Ba}_{n+1}\text{Pb}_n\text{O}_{3n+1}$ (ref. 4), all of which exhibit a trend from semiconducting to metallic behaviour with increasing number of perovskite layers (n). We report here the synthesis of a family of organic-based layered halide perovskites, $(\text{C}_4\text{H}_9\text{NH}_3)_2(\text{CH}_3\text{NH}_3)_{n-1}\text{Sn}_n\text{I}_{3n+1}$, which show a similar transition from semiconducting to metallic behaviour with increasing n . The incorporation of an organic modulation layer between the conducting tin iodide sheets potentially provides greater flexibility for tuning the electrical properties of the perovskite sheets, and we suggest that such an approach will prove valuable for exploring the range of transport properties possible with layered perovskites.

Synthesis of the tin(II) organic-based layered perovskites is performed in an argon atmosphere to prevent oxidation. Unlike the oxide perovskites, where high temperatures are generally required for stabilization, these complex perovskites form near room temperature. First, separate solutions of SnI_2 and $\text{C}_4\text{H}_9\text{NH}_2 \cdot \text{HI} / \text{CH}_3\text{NH}_2 \cdot \text{HI}$ are prepared using a concentrated (57 wt%) aqueous HI solvent at 90 °C. Stoichiometric quantities of SnI_2 , $\text{C}_4\text{H}_9\text{NH}_2 \cdot \text{HI}$ and $\text{CH}_3\text{NH}_2 \cdot \text{HI}$ are used in the preparation of compounds with $n = 1-3$. An excess of $\text{CH}_3\text{NH}_3\text{SnI}_3$ is necessary for a high yield of compounds with $n = 4$ and 5. After mixing the SnI_2 and organic amine solutions, a yellow solution results. Slow cooling at a rate of 2–5 °C h⁻¹ from 90 to –10 °C results in a high yield of plate-like crystals, which are dark red ($n = 1$) or black ($n = 2-5$). For $n = 1-4$, X-ray powder diffraction indicates single-phase samples, whereas for $n = 5$, the products are ~90% pure, with the balance of the crystals consisting of the $n = 4$ phase.

The orthorhombic structure of the $n = 3$ compound, $(\text{C}_4\text{H}_9\text{NH}_3)_2(\text{CH}_3\text{NH}_3)_2\text{Sn}_3\text{I}_{10}$ (Fig. 1, Table 1), is closely related to the $n = 3$ Ruddlesden–Popper phase $\text{Sr}_4\text{Ti}_3\text{O}_{10}$, (ref. 5) with the inorganic SrO 'rock-salt' layer replaced by a double organic layer of n -butylammonium chains. This type of structure is also known for a variety of insulating metal halides, some of which are nearly ideal two-dimensional magnets or have nonlinear optical properties^{6,7}. Within the triple perovskite layer, the 'inner' $\text{CH}_3\text{NH}_3\text{SnI}_3$ perovskite layer resembles the metallic

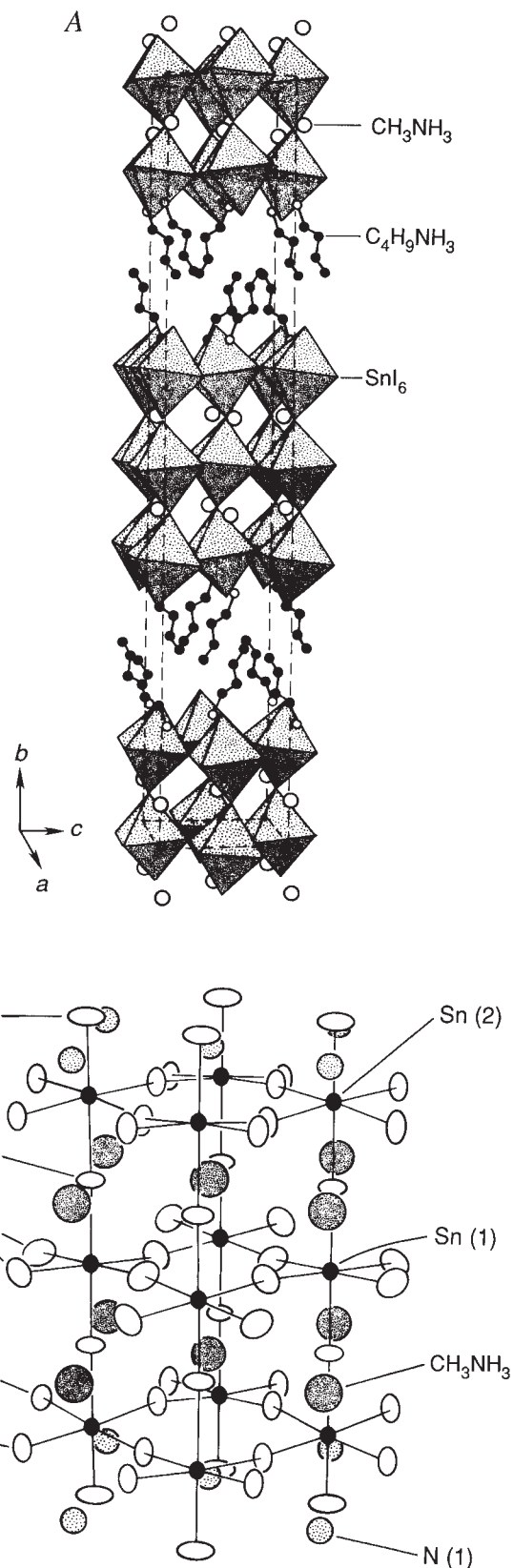


FIG. 1 A, Schematic crystal structure of $(\text{C}_4\text{H}_9\text{NH}_3)_2(\text{CH}_3\text{NH}_3)_2\text{Sn}_3\text{I}_{10}$ ($n = 3$), as determined using a Siemens single-crystal diffractometer and graphite-monochromatized $\text{Mo K}\alpha$ radiation. Data were collected at –50 °C in the ω – 2θ scan mode up to $2\theta = 55^\circ$. θ -dependent empirical absorption corrections based on 10 ψ -scans were applied to all intensity data. The SnI_6 octahedra are shown in a polyhedral representation with the iodine atoms removed for clarity. The carbon atoms of the CH_3NH_3 group (represented by the larger white balls) cannot be fully located with the Fourier difference maps, suggesting significant disorder as indicated by the residual electron density around the N(2) nitrogen. The smaller white balls represent the N(1) nitrogens. Refinement of the

n -butyl fragment also indicates significant rotational disorder resulting in partial occupancy over multiple sites for the terminal carbons. For clarity, only one configuration of the terminal carbons is shown. Crystallographic data are presented in Table 1. B, Detailed structure of the $\text{Sn}-\text{I}$ perovskite trilayers showing atom labelling and anisotropic thermal ellipsoids.

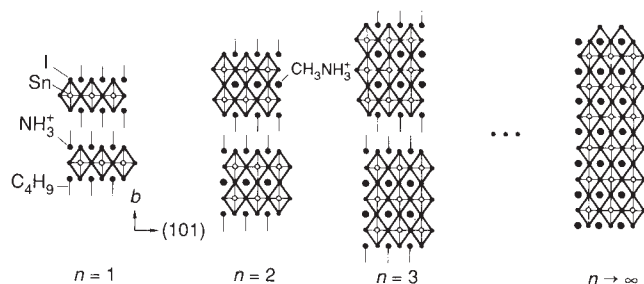


FIG. 2 Schematic representation of the $n=1$ to ∞ (cubic perovskite) compounds. The long unit-cell dimension perpendicular to the perovskite sheets, b , increases in the order 27.576 (2), 39.395 (5), 51.870 (3), 64.29 (2), 76.79 (2) Å at room temperature as n increases from 1 to 5.

cubic perovskite^{8,9}, $\text{CH}_3\text{NH}_3\text{SnI}_3$ ($n=\infty$), with nearly ideal octahedral coordination of iodine about the Sn(1) atoms. The I—Sn(1)—I bond angles are all $90 \pm 1^\circ$, while the Sn(1)—I distances are essentially equal with two 3.123 (3)-Å Sn(1)—I(2) apical bonds and four 3.127 (1)-Å Sn(1)—I(3) equatorial bonds (compared with the six equal 3.120 Å Sn—I bonds in cubic $\text{CH}_3\text{NH}_3\text{SnI}_3$). There is a slight tilting of the SnI_6 octahedra relative to the ideal perovskite, with a bridging Sn(1)—I(3)—Sn(1) angle of $172.8 (2)^\circ$.

The two 'outer' perovskite layers have Sn(2) atoms located within a more distorted octahedral environment with an I(4)—Sn(2)—I(5) angle of $93.03 (2)^\circ$, leading to an alternate tilting of the apical I(5) iodines. Bond distances range from 3.003 (4) Å for the apical Sn(2)—I(5) bond (toward the organic layer) to 3.284 (4) Å for the opposite Sn(2)—I(2) apical bond. The equatorial Sn(2)—I(4) and Sn(2)—I(1) distances are virtually equal at 3.136 (2) and 3.132 (2) Å. The outer perovskite layers are somewhat more buckled than the inner layer, with bridging Sn(2)—I(1,4)—Sn(2) angles of $166.4 (1)^\circ$ and $173.6 (1)^\circ$.

The organic fragments show significant disorder within the structure, with the methylammonium cation apparently rotating or significantly disordered as in the cubic $\text{CH}_3\text{NH}_3\text{PbI}_3$ perovskite structure¹⁰. The nitrogen atoms attached to the butyl chain are situated off the plane formed by the outer apical I(5) iodines and are displaced towards the Sn(2) planes. The bond between the carbon atom and nitrogen is directed nearly perpendicular to the a - c plane, with the rest of the chain extending towards

the gap and showing significant rotational disorder. Van der Waals interactions between the two hydrocarbon layers hold the organic layers together, and create a highly anisotropic structure.

For the structures with $n=1-5$ (Fig. 2), the b -axis (perpendicular to the perovskite sheets) increases by ~ 12 Å for each increment in n as a result of inserting two more $\text{CH}_3\text{NH}_3\text{SnI}_3$ perovskite layers per unit cell. Additionally, the orthorhombic distortion between the a and c lattice parameters decreases with increasing n , and ranges from $a=8.839 (1)$ and $c=8.626 (1)$ Å for $n=1$, to $a=8.840 (1)$ and $c=8.807 (1)$ Å for $n=5$. The average in-plane lattice parameter $(a+c)/2$ increases from 8.732 to 8.824 Å as n increases from 1 to 5. For comparison, $\sqrt{2}a=8.824$ Å for the cubic ($n \rightarrow \infty$) $\text{CH}_3\text{NH}_3\text{SnI}_3$ structure.

Transport measurements, made using a standard four-point geometry on pressed pellets of samples with $n=1-5$ and on the three-dimensional $\text{CH}_3\text{NH}_3\text{SnI}_3$, are shown in Fig. 3. Increasing n results in a marked drop in resistivity and a transition from semiconducting ($n < 3$) to metallic ($n > 3$) behaviour as in the layered oxide perovskites²⁻⁴. For $n=3$, the resistivity decreases

TABLE 1 Crystallographic data for $(\text{C}_4\text{H}_9\text{NH}_3)_2(\text{CH}_3\text{NH}_3)_2\text{SnI}_{10}$

Atom	Position	x	y	z	$U_{\text{eq}}(\text{\AA}^2)$
Sn(1)	4a	0.5	0	0.5	0.0392
Sn(2)	8f	0.5	0.12298 (5)	0.5011 (3)	0.0405
I(1)	8e	0.25	0.11581 (6)	0.75	0.0659
I(2)	8f	0.5	0.05995 (6)	0.4708 (3)	0.0758
I(3)	8e	0.25	0.00380 (7)	0.75	0.0893
I(4)	8e	0.25	0.12637 (6)	0.25	0.0677
I(5)	8f	0.5	0.18013 (6)	0.5529 (3)	0.0916
N(1)	8f	0	0.3364 (7)	0.031 (4)	0.09 (1)*
N(2)	8f	0.5	0.0621 (9)	0.009 (5)	0.14 (2)*
C(1)	8f	0.5	0.191 (1)	0.031 (6)	0.15 (2)*
C(2)	8f	0.5	0.203 (1)	0.144 (6)	0.13 (3)*
C(3)†	16g	0.439 (8)	0.231 (1)	0.120 (6)	0.15 (3)*
C(4)‡	8f	0.5	0.2496	0.1945	—

Orthorhombic cell; $a=8.795 (1)$ Å, $b=51.921 (8)$ Å, $c=8.858 (2)$ Å; space group Cmca , $Z=4$; observed reflections 1,189; $R=0.050$, $R_w=0.060$. ($U_{\text{eq}}=(U_{11}U_{22}U_{33})^{1/3}$.)

* Refined isotropically.

† Occupancy = $\frac{1}{2}$ as a result of orientational disorder of butylammonium cations.

‡ Idealized position based on refined C(1)—C(3) positions.

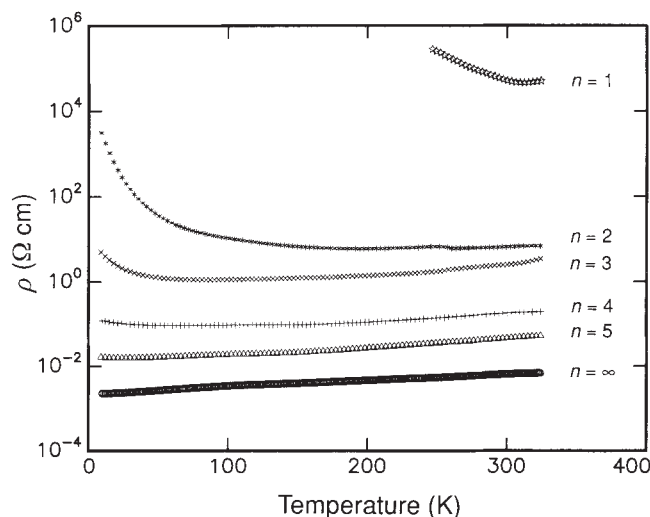


FIG. 3 Resistivity as a function of temperature for pressed-pellet samples of $(\text{C}_4\text{H}_9\text{NH}_3)_2(\text{CH}_3\text{NH}_3)_{n-1}\text{SnI}_{3n+1}$ ($n=1-5$) and the cubic (at room temperature) perovskite ($n \rightarrow \infty$) $\text{CH}_3\text{NH}_3\text{SnI}_3$.

with temperature with an upturn below 75 K, presumably due to localization of carriers or grain boundary effects in the pressed pellet. The materials with $n=4$ and 5 become progressively more metallic, with only a very slight upturn in resistivity below 20 K. Room-temperature Hall effect measurements on pressed pellets of $n=3$ and 5 materials indicate that the carriers are holes. However, even for $n=5$, the Hall carrier concentration is only $7 \times 10^{18} \text{ cm}^{-3}$, approximately two orders of magnitude lower than in the cuprate superconductors, and corresponding to only 0.002 carriers per Sn atom. For comparison, the isotropic $\text{CH}_3\text{NH}_3\text{SnI}_3$ ($n=\infty$) has recently been shown to be a low-carrier-density p-type metal with a well-defined plasma edge and a carrier density between 0.003–0.005 carriers per Sn atom⁸. Attempts to dope these structures and thereby increase the carrier density are presently underway.

To give some demonstration of the structural flexibility provided by the family of organic-based layered perovskites, we have synthesized the more general compounds, $\text{A}_2(\text{CH}_3\text{NH}_3)_{n-1}\text{SnI}_{3n+1}$, where A is $\text{CH}_3(\text{CH}_2)_m\text{NH}_3$ ($m=2-12$) or phenethylammonium (D. B. M., unpublished work). For long-chain substitutions (large m), the structures provide an interesting analogy to lipid bilayers and membranes enabling, for example, the use of solid-state techniques to study conformational changes occurring with the bilayer¹¹. Varying the length of the alkyl chain in the organic layer should also control the degree of coupling between sets of conducting perovskite layers whereas introduction of more complicated organics with unsatu-

rated hydrocarbon chains or rings may introduce interactions with carriers in the perovskite planes, perhaps providing an opportunity for excitonic superconductivity¹² in a two-dimensional system.

It is worth noting that although the full importance of the inorganic modulation layers in the copper oxide perovskites [Cu-O 'chain layer', BiO, TlO and so forth] is still debated, the ability to use these layers to tune the electrical properties of the conducting perovskite sheets, without introducing substantial disorder into these 'active' layers (by analogy with modulation-doped and multilayer semiconductors), undoubtedly plays a significant role in enabling high-temperature superconductivity to appear in these unusual structures. The extra flexibility introduced with the organic modulation layer makes this an interesting family to explore for new superconducting materials, as well as to test existing theories of high-temperature superconductivity that do not rely on having a Cu—O framework, such as the van Hove¹³ and interlayer Coulomb-coupling¹⁴ models. More generally, organic-based layered perovskites may prove useful for tailoring scientifically and technologically interesting conducting materials by a suitable choice of the organic modulation layer and perovskite matrix. □

Received 2 March; accepted 5 May 1994.

1. Bednorz, J. G. & Müller, K. A. *Z. Phys.* **B64**, 189–193 (1986).
2. Mohan Ram, R. A., Ganguly, P. & Rao, C. N. R. *J. Solid St. Chem.* **70**, 82–87 (1987).
3. Mohan Ram, R. A., Ganapathi, L., Ganguly, P. & Rao, C. N. R. *J. Solid St. Chem.* **63**, 139–147 (1986).
4. Cava, R. J. et al. *Phys. Rev.* **B46**, 14101–14104 (1992).
5. Ruddlesden, S. N. & Popper, P. *Acta crystallogr.* **11**, 54–55 (1958).
6. Arend, H., Huber, W., Mischgofsky, F. H. & Richter-Van Leeuwen, G. K. *J. Cryst. Growth* **43**, 213–223 (1978).
7. Calabrese, J. et al. *J. Am. chem. Soc.* **113**, 2328–2330 (1991).
8. Mitzi, D. B., Feild, C. A., Schlesinger, Z. & Laibowitz, R. B. *J. Solid St. Chem.* (in the press).
9. Yamada, Y., Matsui, T., Tsuritani, T., Okuda, T., & Ichiba, S. *Z. Naturf.* **45a**, 307–312 (1990).
10. Xu, Q., Eguchi, T., Nakayama, H., Nakamura, N. & Kishita, M. *Z. Naturf.* **46a**, 240–246 (1990).
11. Needham, G. F., Willett, R. D. & Franzen, H. F. *J. phys. Chem.* **88**, 674–680 (1984).
12. Little, W. A. *Phys. Rev. A* **134**, 1416–1424 (1964).
13. Newns, D. M., Tsuei, C. C., Pattnaik, P. C. & Kane, C. L. *Comments cond. Mater Phys.* **15**, 273–302 (1992).
14. Harshman, D. R. & Mills, A. P. *Jr Phys. Rev.* **B45**, 10684–10712 (1992).

ACKNOWLEDGEMENTS. We thank B. A. Scott and J. Korp for discussion, R. B. Laibowitz for help with the Hall measurements, and M. Ritter for use of a closed-cycle refrigeration system for measuring resistivity. One of us (W.H.) also thanks the R. A. Welch Foundation for support.

Mechanism of the transition from fractal to dendritic growth of surface aggregates

Harald Brune, Christoph Romainczyk, Holger Röder & Klaus Kern

Institut de Physique Expérimentale,
Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne,
Switzerland

THE similarity of many patterns formed in non-equilibrium growth processes in physics, chemistry and biology is conspicuous, and many attempts have been made to discover common mechanisms underlying their formation¹. A central question is what causes some patterns to be dendritic (symmetrically branched, like snowflakes) and others fractal (randomly ramified). In general, the transition from fractal to dendritic growth is regarded as a manifestation of the predominance of anisotropy over random noise in the growth process. In electrochemical deposition, this transition is observed as the growth speed is varied^{2,3}. Here we report a crossover from fractal to dendritic growth in two dimensions on the microscopic scale. We use the scanning tunnelling microscope to study diffusion-limited aggregation of silver atoms on a Pt(111) surface. The

transition occurs as the deposition flux is increased, and our observations suggest that the increasing importance of anisotropy of edge diffusion at higher flux is responsible for this crossover. We anticipate that a similar phenomenon may operate in three-dimensional crystal growth.

The direct experimental analogue to the two-dimensional diffusion limited aggregation (DLA) simulations^{4,5} is vapour-phase epitaxy of metals on single-crystal metal surfaces at low temperatures. The film atoms, once adsorbed on the surface from the gas phase, diffuse in a random walk until they stick irreversibly to the perimeter of a growing aggregate. If perimeter diffusion is prohibited or is negligibly small, classical fractals result⁶. If a certain perimeter mobility is allowed and this mobility is anisotropic, however, deviations from this behaviour can occur, as will be demonstrated here.

In our experiments the parameters controlling growth are the temperature and flux of deposition, the former determining the perimeter mobility and the latter the growth velocity. The experiments were performed in ultra-high vacuum with standard facilities for sample preparation and film evaporation. We applied variable-temperature scanning tunnelling microscopy (STM)⁷ to image the Ag cluster morphology isothermally to deposition because the patterns formed are metastable. In the STM images the derivative $\partial z/\partial x$ of the lines of constant tunnel current has been

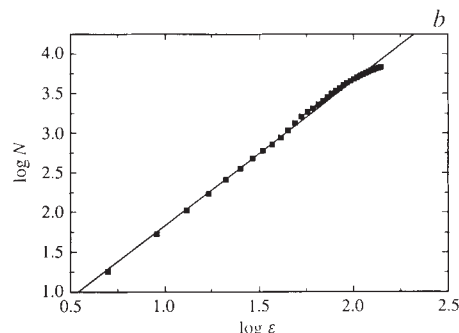
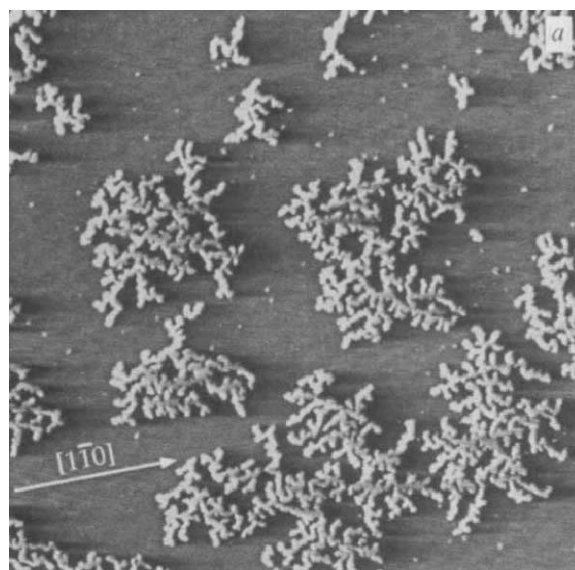


FIG. 1 a, STM image showing fractal Ag aggregates grown on Pt(111) at 110 K and a deposition flux (R) of 1.6×10^{-5} monolayers per second; one monolayer (ML) corresponds to the density of the Pt(111) substrate of 1.50×10^{15} atoms cm^{-2} (size, $1,200 \text{ \AA} \times 1,200 \text{ \AA}$; coverage $\theta = 0.12$ ML). b, Ag cluster (lower right in a) obeys a fractal dimension of $D = 1.78$ for the straight line fitted to the data; ϵ is the border length of boxes in pixel (1 pixel corresponds to $2.3 \text{ \AA}^2$), N is the number of pixels where the surface is covered by Ag.

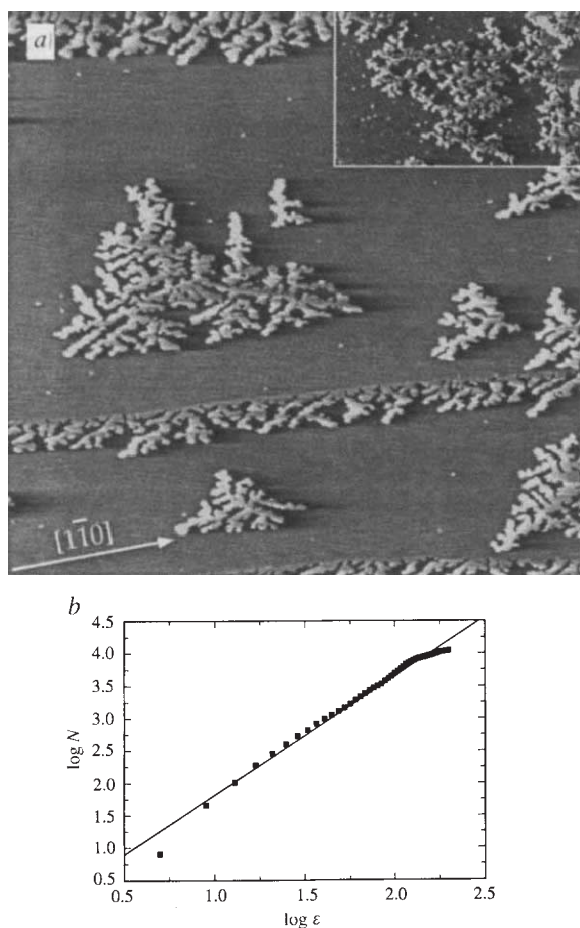


FIG. 2 *a*, STM image showing the dendritic shape of the Ag aggregates grown at 130 K with a flux (R) of $1.1 \times 10^{-3} \text{ ML s}^{-1}$ ($1,200 \times 1,200 \text{ \AA}$, $\theta = 0.12 \text{ ML}$), note also that two Pt substrate steps are crossing the image almost horizontally. *b*, Mass-length scaling of the biggest dendrite in *a*. Inset in *a*, Transition from the dendritic to fractal growth at 130 K on lowering the deposition flux by two orders of magnitude ($830 \times 520 \text{ \AA}$, $\theta = 0.12 \text{ ML}$, $R = 1.6 \times 10^{-5} \text{ ML s}^{-1}$).

recorded. They therefore represent the surface as it appears when illuminated from the left.

The two-dimensional Ag aggregates shown in Fig. 1*a* have been grown at 110 K at a low Ag flux. Large clusters ($\sim 3,000$ atoms) with an open ramified structure are formed under these conditions. The branches of the clusters frequently alter their direction of growth and thus show no long range

correlation with the trigonal substrate symmetry. The branches are of monoatomic height, their thickness is almost constant over the entire aggregate and much smaller than its radius of gyration. In fact the arms are only 2 ± 1 atoms wide as determined from the total arm length and the cluster size. The shape of the Ag aggregates grown at 110 K is very similar to that of fractal aggregates simulated with the classical DLA computer codes either performed off-lattice (no anisotropy)^{5,8} or on-lattice with noise dominating lattice anisotropy^{4,5}.

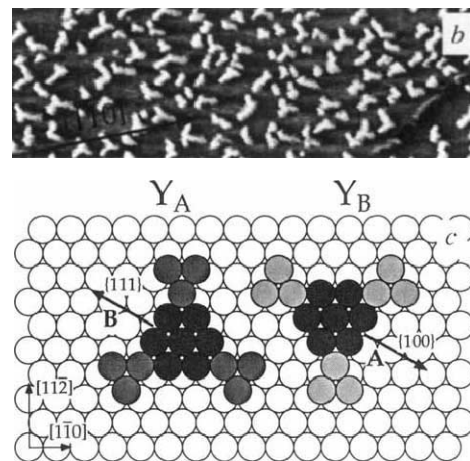
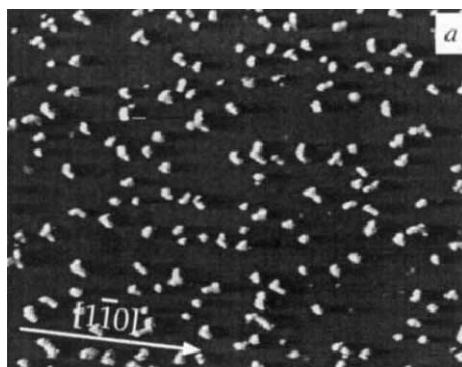
In order to examine the fractal character of the Ag aggregates shown in Fig. 1*a* we have determined their fractal dimension D . For that purpose we arranged boxes with border length ϵ centred at the centre of mass of the aggregate and counted the number of pixels N with grey levels corresponding to Ag as a function of box size ϵ (ref. 1). $N(\epsilon)$ is shown in Fig. 1*b* for the large Ag cluster located on the lower right-hand side of Fig. 1*a*. It is evident that the number of Ag-covered pixels N scales over two orders of magnitude with the box size as ϵ^D , where $D = 1.78$. The scaling is limited at small lengths by the finite apparent branch width of 14 Å (partly due to curvature of the STM tip) and at large lengths by the finite cluster size. By investigating the variation of mass with size of numerous Ag aggregates grown at 110 K, we find the average fractal dimension to be $D = 1.76 \pm 0.07$. This experimental value is in good agreement with two-dimensional DLA simulations^{4,5}.

A drastic change of the aggregate patterns is observed when the flux is increased by two orders of magnitude. Figure 2*a* demonstrates that Ag clusters grown with this higher growth speed obey a nice dendritic pattern. It shows the characteristic 'back-bones', whose orientation is determined by the crystalline anisotropy of the substrate. In snow-flake terminology this dendrite is of the P2a type (plane P, with irregular number of branches 2, three branched a). Although the qualitative growth form has changed dramatically, the fractal dimension is almost unaffected. From the mass-length scaling plot in Fig. 2*b* we infer a fractal dimension of $D = 1.77 \pm 0.05$ for the biggest cluster in Fig. 2*a*, which agrees with that of the randomly ramified aggregates of Fig. 1*a*. This is expected from DLA for the relatively small cluster sizes under consideration here⁹.

It is important to notice that the slight increase in temperature to 130 K chosen for Fig. 2*a* (in order to obtain clusters with comparable size to those in Fig. 1*a*) has no influence on the cluster shape. The inset in Fig. 2*a* shows that randomly ramified patterns (with no preferred orientation) result at 130 K through application of the low flux. On the other hand, dendrites also grow at 110 K upon deposition with the high flux used in Fig. 2*a* (not shown here). Therefore the parameter that drives the crossover from ramified to dendritic patterns is the deposition flux, that is, the growth speed of the aggregate.

These observations are in agreement with earlier work on other systems. Increasing the growth speed (Zn^{2+} concentration

FIG. 3 *a, b*, STM images of small Ag₂ clusters grown on Pt(111) at 80 K with an average cluster size of $n = 13$ in *a* and $n = 21$ atoms in *b* respectively; (*a*), $600 \times 470 \text{ \AA}$, $\theta = 0.037 \text{ ML}$, $R = 3.7 \times 10^{-4} \text{ ML s}^{-1}$. *b*, $600 \times 190 \text{ \AA}$, $\theta = 0.12 \text{ ML}$, $R = 3.6 \times 10^{-3} \text{ ML s}^{-1}$. *c*, Schematic model showing the two possible Y-shaped cluster orientations.



and voltage) in electrochemical deposition resulted in transition from fractal to dendritic patterns^{2,3}. In fluid systems (Saffman–Taylor instability in the propagation of a low-viscosity medium into one of a high viscosity) dendrites have been found for high pressures and anisotropy whereas low expansion rates produced fractals^{10–12}. Side branching, the necessary condition for dendrites, can also be obtained by the presence of a bubble at the interface tip that prevents its splitting and increases its propagation speed¹³. The hydrodynamic experiments were simulated with a statistical mechanic approach that demonstrated the importance of anisotropy for dendritic growth^{14,15}. Similarly in DLA work, this transition from fractal to dendritic patterns is obtained when anisotropy dominates noise^{8,16–18}. The microscopic mechanism that establishes anisotropy only at increased growth rates to the macroscopic pattern shape, however, can be analysed for the present system.

From a comparison with DLA studies, it is found that the anisotropy that characterizes dendrites here comprises preferential growth in only three of the six $\langle 112 \rangle$ directions⁹. The nature of this anisotropy for growth of Ag on Pt(111) is linked to the trigonal symmetry of this surface. A densely packed cluster on a face-centred cubic (111) surface is bound by two types of edges with atomically different structures (A and B edges): A edges are $\{100\}$ facets whereas B edges are $\{111\}$ facets (Fig. 3c). This microscopic difference is partly reflected in a different activation barrier for perimeter diffusion. To identify which of the close-packed Ag-edge orientations is of A-type and which is of B-type, it is important to note that the Ag atoms reside on f.c.c. sites on the Pt(111) surface for coverages of up to a monolayer. This site can be deduced from the perfect (dislocation-free) attachment of Ag islands to the neighbouring Pt layer at ascending steps (stacking faults would produce undulations in the STM images¹⁹, but none is evident in the figures—for example, the Ag islands are flat at the step edges in Fig. 2a). For Ag on Pt(111), our investigations (analogous to those of Michely *et al.*²⁰) show that diffusion along B steps is faster than along A steps below 200 K, and thus that the former has a lower activation barrier. Interestingly, this is the reverse of what is found for homoepitaxial systems²¹.

This kinetic difference is especially pronounced at low temperature. Figure 3a and b shows the initial branching of the cluster seed at 80 K. The smaller islands that appear almost spherical in Fig. 3a are heptamers, and constitute the seed particle. The biggest islands are forming Y's with arms 120° apart. Y-branching of the heptamer is possible in two ways: by adding Ag adatoms to its A-steps or to its B-steps (see left- and right-hand side of Fig. 3c). In our experiment we only observe the Y_A -cluster. This anisotropy in growth is even better seen for somewhat bigger clusters as shown in Fig. 3b (note the different orientation of the crystallographic $[110]$ direction compared to Fig. 3a). The branches of the Y_A s are exactly in the preferred growth directions found for the bigger dendrite at 130 K. Because Ag atoms arrive at random at each edge orientation, the fact that we find only the Y_A type must involve perimeter diffusion along B-steps, whereas diffusion along A-steps is frozen in. There is, however, a second necessary condition for the observed growth of Y_A s, that applies after the addition of the first three atoms to the heptamer (which is most plausibly done at A-steps): the Ag_{10} aggregate is then exclusively bound by B-steps. In order to grow a stable expansion in the A-direction, as observed, two diffusing perimeter atoms have to meet each other at the aggregate's corners.

Such an event, however, is the more probable the more diffusing atoms similarly are present at the island perimeter. Their quantity is directly related to the flux. At the high flux used in Fig. 2a there are on the average 100 atoms arriving per second at the aggregate's perimeter, whereas there is only one per second at the low flux of Fig. 1a. In addition, the mean free path of a diffusing perimeter atom increases only with square root of the diffusion time (one-dimensional random walk). Therefore the

probability for two atoms to meet at a corner is greater at increased flux, and, due to faster migration along B-steps, much greater at those corners that point in the A-direction. Thus dendrites with trigonal growth perpendicular to A-steps are formed for high flux. At low flux, on the other hand, interaction of two diffusing particles at the island corner is much less probable, and predominantly single non-interacting mobile atoms are present at the step. Anisotropy therefore loses its importance in favour of noise, and the aggregate grows randomly. □

Received 30 December 1993; accepted 31 March 1994.

1. Takayasu, H. *Fractals in the Physical Sciences* (Manchester Univ. Press, New York, 1990).
2. Sawada, Y., Dougherty, A. & Gollub, J. P. *Phys. Rev. Lett.* **56**, 1260–1263 (1986).
3. Grier, D., Ben-Jacob, E., Clarke, R. & Sander, L. M. *Phys. Rev. Lett.* **56**, 1264–1267 (1986).
4. Witten, T. A. & Sander, L. M. *Phys. Rev. Lett.* **47**, 1400–1403 (1981).
5. Meakin, P. *Phys. Rev. A* **27**, 1495–1507 (1983).
6. Hwang, R. Q., Schröder, J., Günther, C. & Behm, R. J. *Phys. Rev. Lett.* **67**, 3279–3282 (1991).
7. Röder, H., Hahn, E., Brune, H., Bucher, J. P. & Kern, K. *Nature* **366**, 141–143 (1993).
8. Eckmann, J. P., Meakin, P., Procaccia, I. & Zeitak, R. *Phys. Rev. Lett.* **65**, 52–55 (1990).
9. Meakin, P. *Phys. Rev. A* **33**, 3371–3382 (1986).
10. Ben-Jacob, E., *et al.* *Phys. Rev. Lett.* **55**, 1315–1318 (1985).
11. Buka, A., Kertész, J. & Vicsek, T. *Nature* **323**, 424–425 (1986).
12. Horváth, V., Vicsek, T. & Kertész, J. *Phys. Rev. A* **35**, 2353–2356 (1987).
13. Couder, Y., Cardoso, O., Dupuy, D., Tavernier, P. & Thom, W. *Europhys. Lett.* **2**, 437–443 (1986).
14. Nittmann, J. & Stanley, H. E. *Nature* **321**, 663–668 (1986).
15. Nittmann, J. & Stanley, H. E. *J. Phys. A* **20**, L1185–L1191 (1987).
16. Meakin, P. *Phys. Rev. A* **36**, 332–339 (1987).
17. Vicsek, T. *Fractal Growth Phenomena* (World Scientific, Singapore, 1989).
18. Eckmann, J. P., Meakin, P., Procaccia, I. & Zeitak, R. *Phys. Rev. A* **39**, 3185–3195 (1989).
19. Brune, H., Röder, H., Boragno, C. & Kern, K. *Phys. Rev. B* **49**, 2997 (1994).
20. Michely, T., Hohage, M., Bott, M. & Comsa, G. *Phys. Rev. Lett.* **70**, 3943–3946 (1993).
21. Ehrlich, G. *Surf. Sci.* **246**, 1–12 (1991).

ACKNOWLEDGEMENTS. This work was supported by the Schweizerische Nationalfonds.

Femtosecond solvation dynamics of water

Ralph Jimenez*, Graham R. Fleming*, P. V. Kumar† & M. Maroncelli†

* Department of Chemistry and The James Franck Institute, The University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, USA

† Department of Chemistry, The Pennsylvania State University, State College, Pennsylvania 16802, USA

THE timescale of the response of solvent molecules to electronic rearrangement of solute molecules has a critical influence on the rates of chemical reactions in liquids^{1–10}. In particular, if the solvent cannot adapt quickly enough to this rearrangement as the reactants pass through the transition state, the evolving products may recross the free-energy barrier, reducing the reaction rate. Computer simulations have shown^{11–18} that the response of a solvent to a change in solute charge distribution is strongly bimodal: there is an initial ultrafast response owing to inertial (mainly librational) motions of the solvent molecules, followed by a slow component owing to diffusive motions. Water seems to be by far the 'fastest' solvent studied so far: simulations predict that well over half of the solvation response for atomic solutes is inertial, happening on a timescale of about 20 femtoseconds^{12,13}. The presence of this ultrafast component implies that solvent friction plays an important role in many aqueous charge-transfer processes^{9,10,19–21}. Experimental verification of this prediction has been lacking, however, in part because of the difficulty of obtaining sufficient time resolution. Here we present experimental measurements of the ultrafast solvation dynamics of a coumarin salt in water. When considered in conjunction with computer simulations, our results demonstrate that a solvent response on a timescale faster than 50 fs can dominate aqueous solvation dynamics.

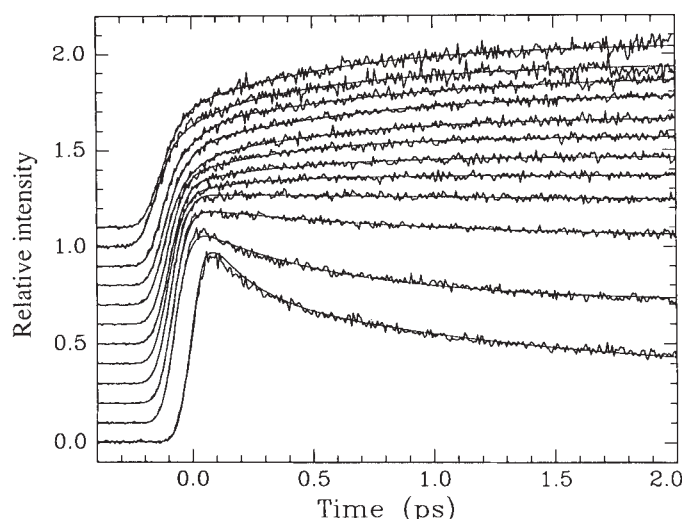


FIG. 1 Femtosecond fluorescence up-conversion data for coumarin 343 anion (sodium salt, 10^{-4} M) in aqueous solution. The traces show data obtained at wavelengths ranging from 460 nm (bottom) to 570 nm (top) in steps of 10 nm.

METHODS. A coherent Mira mode-locked Ti:sapphire laser operating at 850 nm was used. The second harmonic (425 nm) was generated in a 0.4 mm BBO (β -barium borate) crystal and used to excite the coumarin dye, which flowed through a 1 mm quartz cell. The remaining 850-nm light was used to gate the fluorescence by mixing the fluorescence emission in another 0.4 mm BBO crystal. The cross-correlation of the 425 nm and 850 nm pulses was measured as 100–110 fs (full width at half maximum). The intensity of the sum frequency was measured as a function of time delay between the 425 nm and 850 nm pulses, controlled with a stepper motor. By angle tuning the mixing crystal, 12 emission decays were collected at 10-nm intervals. The data plotted here are the relative intensities of upconverted light as a function of delay time. These data were fitted using global analysis, thereby constructing a set of time-resolved emission spectra which reproduce the measured data. The solvation response function $S(t)$ (equation (1)) was then calculated from the set of spectra.

The evolution of the fluorescence spectrum of a suitable probe molecule, following excitation with an ultrashort optical pulse, can be used to monitor polar solvation dynamics by way of the Stokes shift response function^{5–7}

$$S(t) = \frac{\nu(t) - \nu(\infty)}{\nu(0) - \nu(\infty)} \quad (1)$$

Here $\nu(x)$ represents the 'fluorescence frequency' (for instance the peak of the fluorescence spectrum) at times t , zero and infinity. This observed response function can be compared to the time correlation function,

$$C(t) = \frac{\langle \delta\Delta E(0) \delta\Delta E(t) \rangle}{\langle \delta\Delta E^2 \rangle} \quad (2)$$

derived from computer simulation^{12–16}. In this equation ΔE represents the difference in solute-solvent interaction energies between ground and excited states and $\delta\Delta E$ denotes a fluctuation from the mean value ($\delta\Delta E = \Delta E - \langle \Delta E \rangle$). This correlation function measures the loss of memory of random fluctuations in the energy difference ΔE and it can be equated to $S(t)$ provided that linear response holds^{12–16}.

An earlier study²² using the present approach revealed a sub-picosecond response but lacked sufficient time resolution to detect an inertial contribution. In order to achieve the necessary time resolution, we recorded the fluorescence decay at various wavelengths across the emission spectrum using the fluorescence

up-conversion technique. In this method the sample is excited with an ultrashort laser pulse and the resulting fluorescence is resolved by a time-variable gate based on frequency summing the fluorescence with a second ultrashort laser pulse in a nonlinear crystal. Figure 1 shows the fluorescence data collected for coumarin 343 (C343) dissolved in water. The continuous change from rapid decay at the blue end of the spectrum, through a flat response, to rising behaviour at the red end of the spectrum is completely consistent with a continuous shift of the emission spectrum to the red as solvent relaxation proceeds. Thus by reconstructing spectra from these data we can determine how the peak frequency changes with time. The solid lines through the data in Fig. 1 result from a global fitting procedure²³ in which the evolution of the fluorescence spectrum is assumed to reflect an $S(t)$ of the form

$$S(t) = a_g \exp(-\frac{1}{2}\omega_g^2 t^2) + a_1 \exp(-t/\tau_1) + a_2 \exp(-t/\tau_2) \quad (3)$$

Clearly the whole data set is well fitted by this model function for appropriate values of the parameters a_g , ω_g , a_1 , a_2 , τ_1 and τ_2 . The $S(t)$ so obtained is shown in Fig. 2 (dashed line, labelled expt.).

Figure 2 also shows the results of classical molecular dynamics simulations of the $C(t)$ functions (equation (2)) of a small atomic solute (solid line, labelled S^0) and a model of coumarin 343 in water (solid line, labelled Δq). The time correlation function $C(t)$ is calculated from the simulated fluctuations in the energy difference between ground and excited states when the solute is in the ground state. The most striking aspect of both the experimental and simulated results is the large amplitude and ultrafast nature of the initial portion of the response. The molecular basis of the ultrafast component has been the object of much recent discussion^{12,15,24–27}. Simulation makes it clear that librational (rotational) motions of the water molecules are responsible for the ultrafast relaxation^{12,18}. Due to the mass of the oxygen atom, the contribution of translational motions to the initial dynamics is very small. The picture that has emerged from simulation and theory is that the shortest-time behaviour can be viewed as arising from inertial motion of each solvent molecule acting independently of all the others^{14,15,24–27}. Each water molecule commences its motion from a position determined by the liquid structure (that is, its interactions with all the other water molecules and the solute); however, its initial influence is through independent pairwise interactions with the solute. As time proceeds the interactions between water molecules become progressively more important. They lead to the collective oscillations seen in the simulated response and also give the diffusive tail its characteristic timescale. We note that the oscillatory behaviour is quite pronounced in the simulated response of water to small solutes (Fig. 2 and ref. 12) but it is considerably reduced in the solvation response simulated for the experimental probe C343 or similar molecules. Such oscillations are neglected in the global fitting expression and, if present, would probably be undetectable with our current experimental capabilities.

In view of the solvation timescale predicted by simulation it is important to assess the adequacy of the experimental time resolution. As a means of deciding whether the experiment has captured the full extent of the solvation process, we use the time-zero emission spectrum, that is, the emission spectrum of the dye before solvent relaxation. Fee and Maroncelli²⁸ have introduced a technique for estimating the position of the time-zero emission spectrum of coumarin dyes in solution. In this case, the time-zero peak frequency may be estimated by measuring the absorption and emission spectra of C343 in hexane, calculating the difference between the peak values, and equating this value to the difference between the absorption and time-zero spectra in water. By this means, the time-zero spectrum is estimated to peak at $22,222 \text{ cm}^{-1}$. Thus, our estimate for the total Stokes shift, $\nu(0) - \nu(\infty)$, is $1,953 \text{ cm}^{-1}$. Our computer simula-

tion, under the assumption of linear response, gives a predicted shift of $1,850\text{ cm}^{-1}$. From the spectral evolution in our fitting procedure, we observe $1,481\text{ cm}^{-1}$ of the shift, which is $\sim 75\%$ of the estimated total. Thus, we have indeed 'seen' most of the fast solvation response in water. The precise timescale of the initial decay is not well determined in our fits. Tests of the sensitivity of the global fitting procedure indicate that the initial decay is faster than 55 fs and constitutes $>50\%$ of the total decay. Thus, at this stage we cannot meaningfully discuss the precise timescales of the initial experimental and simulated decays and consider instead the overall features of both.

The remarkable agreement between the overall features of the simulation and experiment raises a number of interesting questions. Polarizability is neglected in the model of water used in these simulations. This leads to the absence of the 200 cm^{-1}

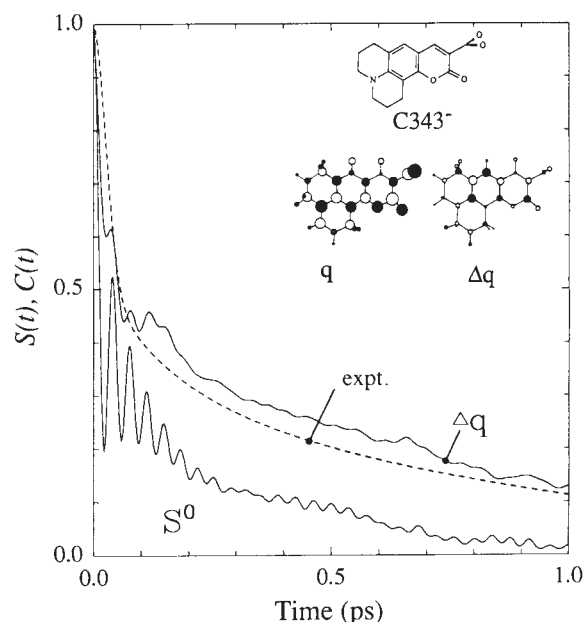


FIG. 2 Experimental ('expt.': $S(t)$, equation (1)) and simulated (' Δq ': $C(t)$, equation (2)) solvation response functions for C343 in water. Also shown is a simulation for a neutral atomic solute with the Lennard-Jones parameters of the water oxygen atom (S^0). The experimental trace is fitted by equation (3) (using the constraint that the long time spectrum match the steady state fluorescence spectrum) as a gaussian component (frequency 38.5 ps^{-1} , 48% of total amplitude) and a sum of two exponential components: 126 fs (20%) and 880 fs (35%). The exponential components agree well with ref. 22. As discussed in the text we estimate that the experiment observes at least 75% of the total spectral shift. Equilibrium classical mechanical simulations¹² were carried out using the SPC/E water potential³⁰. Simulations of atomic ions using other water potentials (for example, ST2, TIP3) produce very similar results. Standard Lennard-Jones parameters were used for the solute. Atom-centred charges were obtained for the coumarin 343 anion from electrostatic potential fits to semi-empirical MNDO wavefunctions calculated using the AMPAC program (Semichem Inc.). The calculated ground state (q) charge distribution and the change in charge distribution on optical excitation (Δq) are shown in the inset. The volume of each sphere is proportional to the charge on that atom and black denotes negative charge. Simulations were run for the ground-state charge distribution and (relevant to the experiment) the fluctuations in the energy difference were calculated using the difference charge distribution between ground and excited states. MNDO is not likely to provide highly accurate excited-state charge distributions, but fortunately the simulated dynamics are not highly sensitive to the details of the excited-state charge distribution.

band in the infrared spectrum which results from dipole-induced dipole interactions²⁹ as well as responses at much higher frequency resulting from intramolecular vibrations¹⁷. However, the translational motions that give rise to the induced dipole moment are present in the simulation, and are felt by the solute molecule. Thus the vibrational motions neglected in our model seem the most likely source of the 'missing' portion of the response in the simulation.

The differences between the dynamics simulated for the atomic and molecular solutes in Fig. 2 are also seen in other solvents such as methanol and acetonitrile²³. We find that these differences reflect the fact that the charge distribution (q), and its change upon electronic excitation (Δq), alternates around the aromatic ring system in typical probe solutes. As a result, the electrostatic field of the solute, to which the solvent responds, has more rapid spatial variation than that of a single charge. Therefore the reaction of the solvent is less collective and hence slower than the response to a simple charge jump (P.V.K. and M.M., manuscript in preparation). But even for molecular solutes such as C343, the inertial component in water still accounts for over half the solvation response.

The prominent ultrafast component observed in both atomic and molecular solutes should have a profound effect on aqueous reaction dynamics. For example the ultrafast component is probably responsible for the greatly enhanced rates of electron transfer between metallocenes in aqueous solutions relative to other polar solvents⁹. Theoretical studies have also emphasized the importance of the ultrafast solvation component for S_N2 reactions in water², and for calculation of deviations of electron transfer rates from Marcus theory as a result of barrier recrossings¹⁹ and quantum dynamical effects^{20,21}. Recrossing of the free energy barrier is influenced by the dynamics of the solvent response: if the solvent is able to respond to the change in charge distribution as the system crosses the transition state, it may be able to stabilize the product and increase the overall reaction rate relative to the case of a 'static' solvent. Our results suggest that solvents are able to respond to changes in reactant configuration on timescales relevant to reactive barrier crossing. □

Received 22 December 1993; accepted 19 April 1994.

- Hynes, J. T. in *Ultrafast Dynamics of Chemical Systems* (ed. Simon, J. D.) 345–381 (Kluwer, Dordrecht, 1994).
- Gertner, B. J., Whittell, R. M., Wilson, K. R. & Hynes, J. T. *J. Am. chem. Soc.* **113**, 74–87 (1991).
- Berne, B. J., Borkovec, M. & Straub, J. *J. phys. chem.* **92**, 3711–3725 (1988).
- Frauenfelder, H. & Wolynes, P. G. *Science* **229**, 337–345 (1985).
- Maroncelli, M., MacInnis, J. & Fleming, G. R. *Science* **243**, 1674–1681 (1989).
- Barbara, P. F. & Jarzaba, W. *Adv. Photochem.* **15**, 1–68 (1990).
- Maroncelli, M. *J. molec. Liq.* **57**, 1–37 (1993).
- Bagchi, B. & Chandra, A. *Adv. chem. Phys.* **80**, 1–126 (1991).
- Weaver, M. J. *Chem. Rev.* **92**, 463–480 (1992).
- Heitele, H. *Angew. Chem. Int. Edn engl.* **32**, 359–377 (1993).
- Impey, R. W., Madden, P. A. & McDonald, I. R. *Molec. Phys.* **48**, 513–539 (1982).
- Maroncelli, M. & Fleming, G. R. *J. chem. Phys.* **89**, 5044–5069 (1988).
- Bader, J. S. & Chandler, D. *Chem. Phys. Lett.* **157**, 501–504 (1989).
- Maroncelli, M. *J. chem. Phys.* **94**, 2084–2103 (1991).
- Carter, E. A. & Hynes, J. T. *J. chem. Phys.* **94**, 5961–5979 (1991).
- Fonseca, P. L. & Ladanyi, B. M. *J. phys. Chem.* **95**, 2116–2119 (1991).
- Muñoz, P. L. & Callis, P. R. *J. chem. Phys.* **100**, 4093–4109 (1994).
- Barnett, R. B., Landman, U. & Nitzan, A. *J. chem. Phys.* **90**, 4413–4422 (1989).
- Smith, B. B., Staib, A. & Hynes, J. T. *Chem. Phys.* **176**, 521–537 (1993).
- Bader, J. S., Kuharski, R. A. & Chandler, D. *J. chem. Phys.* **93**, 230–236 (1990).
- Song, X. & Marcus, R. A. *J. chem. Phys.* **99**, 7768–7773 (1993).
- Jarzaba, W., Walker, G. C., Johnson, A. E., Kahlow, M. & Barbara, P. F. *J. phys. Chem.* **92**, 7039–7041 (1988).
- Rosenthal, S. J., Jimenez, R., Fleming, G. R., Kumar, P. V. & Maroncelli, M. *J. molec. Liq.* (in the press).
- Maroncelli, M., Kumar, P. V. & Papazyan, A. *J. phys. Chem.* **97**, 13–17 (1993).
- Roy, S. & Bagchi, B. *J. chem. Phys.* **99**, 9938–9943 (1993).
- Raineri, F. O., Resat, H., Perng, B.-C., Hirata, F. & Friedman, H. L. *J. chem. Phys.* **100**, 1477–1491 (1994).
- Stratt, R. & Cho, M. *J. chem. Phys.* **100**, 6700–6708 (1994).
- Fee, R. S. & Maroncelli, M. *Chem. Phys.* (in the press).
- Guillot, B. *J. chem. Phys.* **95**, 1543–1551 (1991).
- Beredsen, H. J. C., Grigera, J. R. & Straatsma, T. P. *J. phys. Chem.* **91**, 6269–6271 (1987).

ACKNOWLEDGEMENTS. We thank R. Inaba for help with the experiment. R.J. acknowledges the National Physical Science Consortium for a graduate fellowship. This work was supported by the US NSF and the Office of Basic Energy Sciences of the US Department of Energy.

Seismic evidence for silicate melt atop the 410-km mantle discontinuity

J. Revenaugh* & S. A. Sipkin†

* Institute of Tectonics, University of California at Santa Cruz, Santa Cruz, California 95064, USA

† US Geological Survey, Golden, Colorado 80401, USA

LABORATORY results demonstrating that basalt to ultrabasic melts become denser than olivine-rich mantle at pressures above 6 GPa (refs 1–3) have important implications for basalt petrogenesis, mantle differentiation and the storage of volatiles deep in the Earth. A density cross-over between melt and solid in the extensively molten Archaean mantle has been inferred from komatiitic volcanism^{4–6} and major-element mass balances⁷, but present-day evidence of dense melt below the seismic low-velocity zone is lacking. Here we present mantle shear-wave impedance profiles obtained from multiple-ScS reverberation mapping for corridors connecting western Pacific subduction zone earthquakes with digital seismograph stations in eastern China, imaging a ~5.8% impedance decrease roughly 330 km beneath the Sea of Japan, Yellow Sea and easternmost Asia. We propose that this represents the upper surface of a layer of negatively buoyant melt lying on top of the olivine → β -phase transition (the 410-km seismic discontinuity). Volatile-rich fluids expelled from the partial melt zone as it freezes may migrate upwards, acting as metasomatic agents^{8,9} and perhaps as the deep 'proto-source' of kimberlites^{10,11}. The remaining, dense, crystalline fraction would then concentrate above 410 km, producing a garnet-rich layer that may flush into the transition zone.

Structural contractions, coordination changes and topological rearrangements in silicate liquids at high pressures and temperatures result in high compressibility and belie the notion that silicate melts are always buoyant^{3,12,13}. In fact, olivine and orthopyroxene float in basic melt at pressures above 6 GPa, with important implications for early mantle differentiation, and raising the possibility that enriched silicate melts might reside stably at mid-mantle depths today. The question is, how can we detect and map mantle melt? The ancient signature (komatiitic volcanism) is largely extinct, deep-mantle xenoliths are rare, and kimberlite coverage is spotty at best, demanding that we adopt some sort of remote-sensing approach. Of the candidate seismological techniques, multiple-ScS reverberations¹⁴ (Fig. 1) combine short horizontal path lengths and high sensitivity to radial stratification, offering an excellent means of dowsing for mantle melt.

We have collected 24 seismograms of transversely polarized (SH) shear waves from intermediate and deep-focus earthquakes in the Kuril, Hokkaido and Izu–Bonin subduction zones, digitally recorded by stations of the Chinese Digital Seismograph Network (CDSN), which sample extensively the mantle wedge beneath the Sea of Japan and easternmost Asia (Fig. 2). The records were divided into seven seismic corridors, deconvolved, low-pass filtered (40-mHz corner frequency), and inverted for corridor-averaged crustal structure and whole-mantle travel time and attenuation (Q_{ScS})^{14,15}. Subtracting the model-predicted synthetic waveforms of the zeroth-order reverberations (ScS_n and sScS_n ; Fig. 1a, b) leaves behind a ~50-min window, following the Love wave-train and terminated by the diffracted major-arc phase SSS_{diff} , almost entirely comprised of first- and higher-order reverberations (Fig. 1a, c). These phases are a rich source of information regarding mantle impedance layering. When this information is put into a migration-like estimator, the results are low-passed profiles of mantle shear-wave reflectivity capable of resolving abrupt impedance variations of 1% or less^{16,17}. The

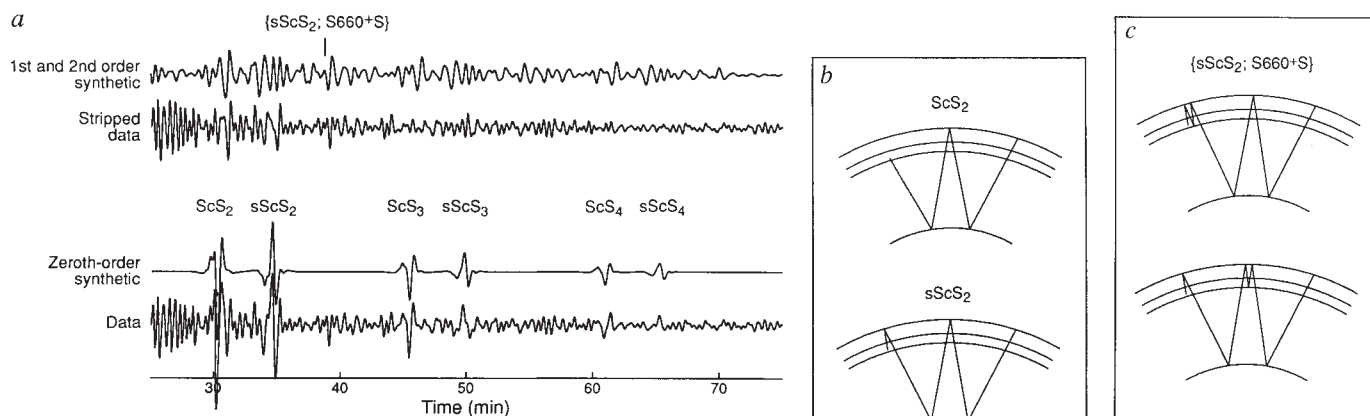


FIG. 1 a, The lower pair of traces compares the deconvolved, low-pass filtered SH-polarized seismogram of the 12 May 1990 Sakhalin Island earthquake (depth, 607 km; $M_b = 6.5$) recorded at station KML of the Chinese Digital Seismograph Network (trace labelled 'data') with the best-fitting model of multiple ScS and sScS phases obtained by waveform inversion¹⁹ ('zeroth order synthetic'). The upper pair of traces compares the stripped data (residual trace) with the model of first- and second-order reverberations corresponding to the discontinuity set of corridor 6 in Fig. 3b. The excellent agreement in energy level and, to a lesser degree, individual waveform packets, testifies to the dominance of deterministic reverberation energy in the reverberative interval and our ability to model it. (The reverberative interval is the portion of seismogram shown, following the Love wave-train and terminating with the first arrivals from the major-arc arrivals, that is, the first arriving waves to have traversed the longer of the two great-circle paths connecting the source and receiver). b, Illustrative ray-paths of the zeroth-order reverberations ScS_2 and sScS_2 . (The subscript "2" refers to the number of core–mantle boundary reflections the wave undergoes.) Multiple ScS

reverberations are ordered by the number of internal reflections (reflections from mantle discontinuities) they experience. ScS_n and sScS_n are zeroth-order reverberations as they only reflect from the mantle's external boundaries. c, Ray-paths of the multiple ScS reverberations forming the first-order reverberation family $\{\text{ScS}_2; \text{S660}^+\text{S}\}$, where each is the combination of two ScS legs (ScS_2 in the bracket notation) and a single topside reflection from the 660-km discontinuity ($\text{S660}^+\text{S}$). In a laterally homogeneous Earth the three arrive in phase, summing constructively to build a single waveform whose amplitude is controlled by the reflection coefficient of the reflector. Migration maps first-order reverberations, such as these, from the amplitude and travel time domains to reflectivity and depth in the mantle.

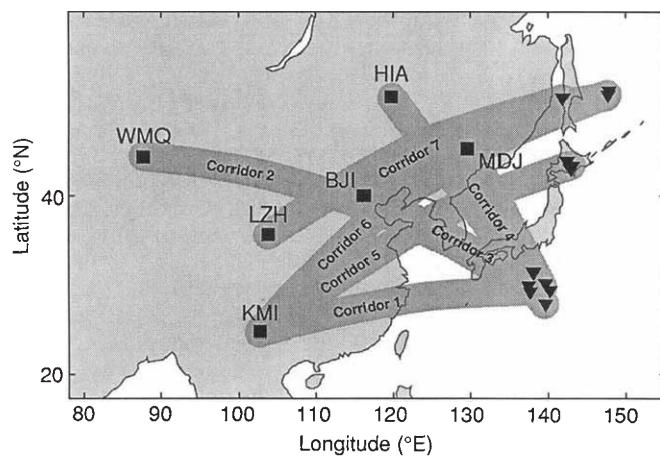


FIG. 2 Mercator projection indicating the intermediate and deep-focus earthquakes (solid triangles), stations of the Chinese Digital Seismograph Network (CDSN) (solid squares) and seismic corridors (mantle swaths connecting source region/receiver pairs) employed in this study. The coverage of discontinuity bounce points provided by first-order reverberations along the corridors is quite dense. Given the dimension of the Fresnel zone for first-order reverberations, which is, roughly speaking, the size of the Korean peninsula (~ 400 km), along-path coverage can be considered continuous.

upper 900 km of the reflectivity profiles for the seven corridors of Fig. 2 are shown in Fig. 3a, b.

All of the corridors have a conspicuous negative reflectivity peak ~ 54 –98 km above the 410-km discontinuity (W in Fig. 3a, b). The peak is also seen in two additional corridors connecting events in Hokkaido and Izu–Bonin with station LZH, but these corridors are not included in the data set because of low-profile signal-to-noise ratios. As no similar feature is seen above

the 660-km discontinuity, W is neither a wave-form mismatch nor processing artefact; nor is it noise-induced, as it appears on all seven corridors crossing a geographically limited region (Fig. 2). (Although some individual corridors near the study region, analysed in previous ScS reverberation studies^{14,17}, also show evidence for the W reflector, they were not modelled with W as they were too few in number to justify the additional reflector.) It is also not well modelled as side-lobe interference of the 410-km discontinuity with a shallower impedance increase, such as the X or L discontinuities of J.R. and Jordan¹⁷ (Fig. 3c). Because of the approximate upper/lower mantle ambiguity inherent in the migration^{14,16}, W may be interpreted as an impedance increase ~ 600 km above the core–mantle boundary. We discount this interpretation on several grounds: (1) W is consistently better modelled by an upper-mantle discontinuity; (2) no lowermost-mantle discontinuity has been observed greater than ~ 350 km from the core–mantle boundary; and (3) the observed depth and impedance variations would be difficult to reconcile with the much decreased areal extent of sampling for a very deep discontinuity.

As it is unlikely that a phase transition would produce an impedance decrease at these depths, we propose an abruptly terminated (<40 km), sill-like accumulation of (potentially) volatile-rich, silicate melt. One alternative is to ascribe W to a compositional boundary, perhaps the upper horizon of a seismically slow, clinopyroxene-rich layer¹⁸. But composition alone cannot account for the correlation, significant at the 87% confidence level, between Q_{scs}^1 (as reported in Sipkin and Revenaugh¹⁹), and layer-W severity, defined as $S_W = (R_{410} - R_W)(z_{410} - z_W)$, where R denotes the SH reflection coefficient and z the path-averaged depth of the subscripted discontinuity. The implication of enhanced attenuation in layer W strongly favours the presence of a melt phase, one whose effect on upper-mantle anelasticity rivals (or exceeds) that of the shallow low-velocity zone observed along six of the seven paths. Although little is known about the elastic properties of silicate melt at these high pressures, there is

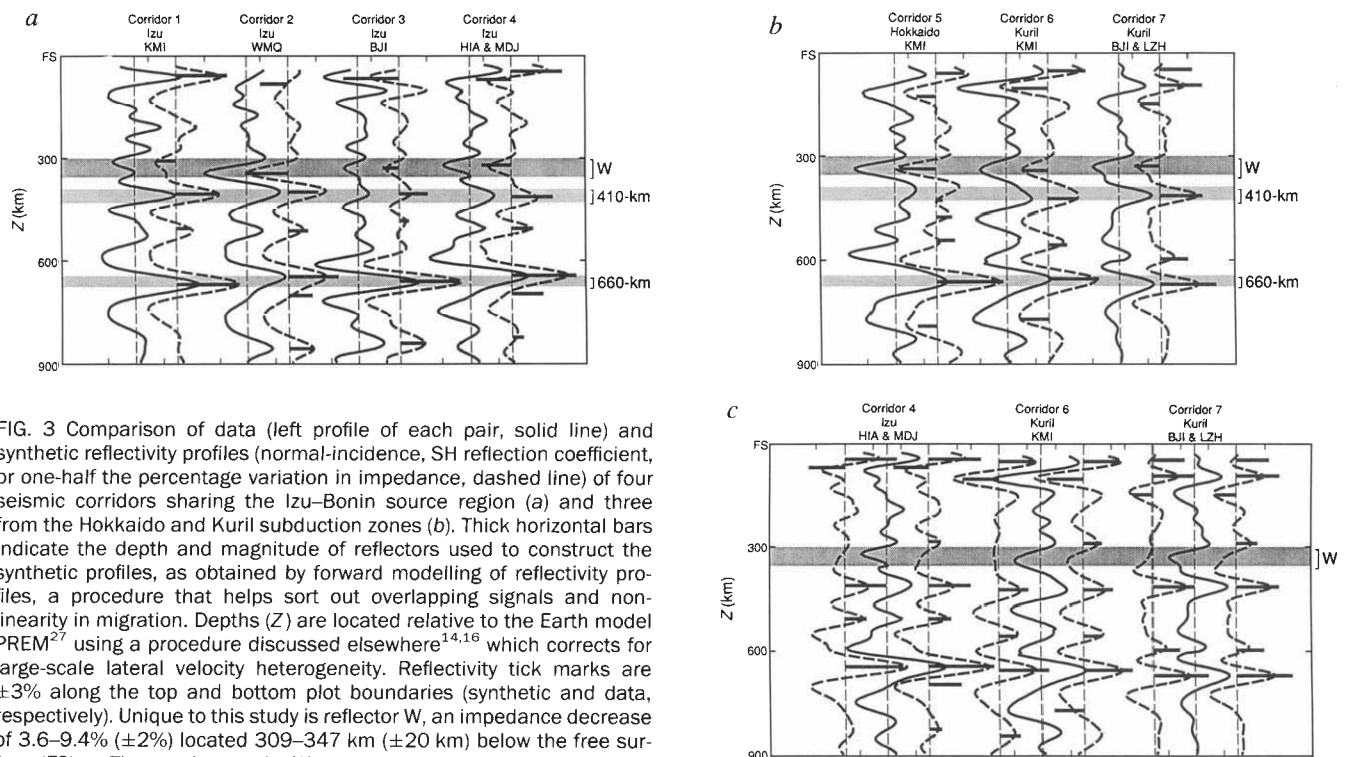


FIG. 3 Comparison of data (left profile of each pair, solid line) and synthetic reflectivity profiles (normal-incidence, SH reflection coefficient, or one-half the percentage variation in impedance, dashed line) of four seismic corridors sharing the Izu–Bonin source region (a) and three from the Hokkaido and Kuril subduction zones (b). Thick horizontal bars indicate the depth and magnitude of reflectors used to construct the synthetic profiles, as obtained by forward modelling of reflectivity profiles, a procedure that helps sort out overlapping signals and non-linearity in migration. Depths (Z) are located relative to the Earth model PREM²⁷ using a procedure discussed elsewhere^{14,16} which corrects for large-scale lateral velocity heterogeneity. Reflectivity tick marks are $\pm 3\%$ along the top and bottom plot boundaries (synthetic and data, respectively). Unique to this study is reflector W, an impedance decrease of 3.6–9.4% ($\pm 2\%$) located 309–347 km (± 20 km) below the free surface (FS). c, The need to model W as an impedance decrease is made clear by comparison of data (centre profile, solid line, of each triplet) with synthetic reflectivity profiles calculated without W (left) and profiles where W is formed through side-lobe interference between the 410-km

discontinuity and a shallower impedance increase (right), neither of which offers a convincing fit to data. The source region and station names of profile corridors are listed above each pair or triplet.

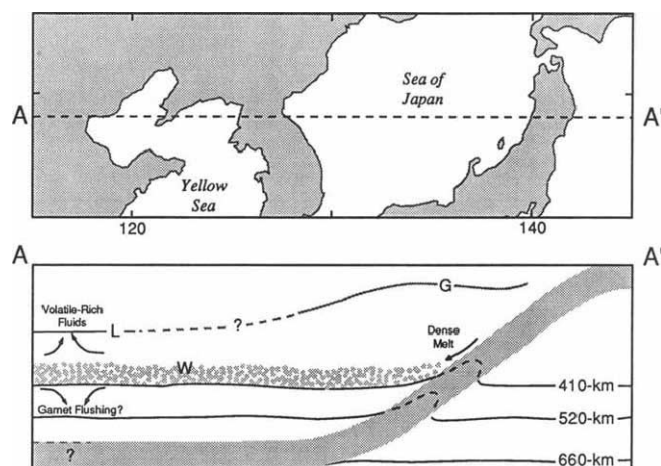


FIG. 4 Cross-section A–A' (bottom box), extending from the westernmost Pacific to eastern China (dashed line of top box), is a schematic depiction of our interpretation of W, namely a sill-like accumulation of negatively-buoyant, potentially volatile-rich, basic to ultrabasic melt supported by the density increase accompanying the olivine $\rightarrow$ β -phase (410-km) transition and protected from general mantle circulation by the horizontally deflected slab imaged by van der Hilst *et al.*²¹. Melt may be produced below ~ 6 GPa (185 km; the depth at which a basic melt becomes denser than the crystalline residue), or transported downward by entrainment in slab flow. The geographic extent of W requires $\sim 2 \times 10^4$ km of melt for 0.5% melt fraction. Eventual freezing of W mantle could trigger ascent of volatile-rich fluids which solidify gradually as they rise, producing a potentially extensive metasomatic layer. Subsequent remelting owing to plume interaction or extensional decompression, aided by high H_2O and CO_2 content, has the potential to produce kimberlite, lamproite and carbonatite magmas³². Over time, the dense crystalline fraction forms a garnet-rich layer which may flush into the transition zone. Downward segregation of the basaltic component has important implications for mantle-wedge depletion and tectosphere growth⁸. Above W, reflector G marks the top of the oceanic low-velocity zone, which fades under the continental margin, and is replaced by L (a shear wave impedance increase) beneath the continent. A similar transition, minus W, is observed beneath northern Australia^{14,17}.

no doubt that it will attenuate seismic waves. Disallowing melt while attributing R_W to higher-than-normal temperatures requires an increase of $\sim 400^\circ\text{C}$ at W (refs 16, 20), resulting in a nearly 40-km depression of the 410-km discontinuity—no more than 5 km of which is seen—unless the anomaly can somehow be confined to the upper reaches of the thin layer separating the 410-km and W discontinuities. Even if this were possible, 400°C is sufficient to produce melt at these depths, returning us to our preferred interpretation.

Assuming equal melt and matrix density, mean R_W (-2.9%) requires a composite 5.8% reduction in shear velocity (v_s)—an underestimate as melt density probably exceeds that of the crystalline matrix and W need not exist over the entire region sampled by first-order reverberations. It is, nonetheless, large enough to be imaged by regional tomography, in fact a 0.3–2% drop in P-wave velocity (v_p) at 340 km beneath the western Sea of Japan and easternmost Asia (well-resolved and falling within a depth interval marked by very little other structure west of the Pacific rim subduction zones) is seen in the tomographic model of van der Hilst *et al.*²¹. Combining tomographic and reverberation results yields an estimate of $\delta v_s/\delta v_p$ similar to the low-velocity zone²², further endorsing a partial melt zone. Unfortunately both v_s and v_p are subject to considerable uncertainty, such that (aggravated by our ignorance of melt viscosity) $\delta v_s/\delta v_p$ is incapable of constraining melt fraction below 1%, assuming that the fluid occurs as a thin film between grains²³ or below 2% if the fluid occurs as tubes along grain boundaries^{24,25}. These upper bounds are capable of producing much greater shear modulus reduction.

The origin of melt and the nature of its radial boundaries are debatable, and we offer several alternatives. (1) The model of van der Hilst *et al.* details a horizontally deflected slab beneath discontinuity W, such that slab devolatilization products (or volatile-catalysed melt) could rise through the transition zone and pond beneath the continental mechanical boundary layer⁸. This is considered unlikely, as it would require an unusually thick mechanical boundary layer beneath the continental margin, and cannot account for W beneath the Sea of Japan. (2) Basic to ultrabasic melts produced below the density cross-over (~ 185 km), or entrained by the slab past that point²⁶, accumulate atop the composite $\sim 5\%$ density increase of the olivine $\rightarrow$ β -phase transition at 410 km (ref. 27); at this point they are gravitationally trapped, and flow outward (Fig. 4). They do not freeze during transport because the appropriate 'wet' adiabat remains within the melt zone over a wide depth range⁶. Complicating this picture is uncertainty in the depth of melt origin and the extent to which slab devolatilization is implicated. The fact that subduction zones are cool focuses attention on the catalytic role of volatiles in the melt-generation process and their effect on melt compressibility. Unfortunately, virtually nothing is known about volatile-charged melt density and compressibility at mid-mantle temperatures and pressures, forcing us to forego further speculation regarding melt production.

Irrespective of source depth, support of melt by the 410-km discontinuity appears likely on theoretical^{1,28} and observational (correlation of S_W and Q_{Scs}^1) grounds, such that discontinuity W marks the upper limit of a ~ 80 -km-thick accumulation of partial melt, whose geographic extent requires a minimum of 2×10^4 km³ of melt (assuming 0.5% average melt content). Geographically, W's position ~ 330 km beneath a continental margin leads us to speculate that some volatile-rich fluids, expelled as melt below discontinuity W freezes, could percolate into the continental tectosphere, serving as potent metasomatic agents and the source of kimberlitic volcanism. At the same time, density-driven flushing of the crystalline fraction could produce a transition zone enriched in garnet, the liquidus phase. The possibility of extensive basalt depletion without surface expression vastly increases the volume of material available for continental tectosphere production⁸, and complicates chemical mass-balance models of crust and mantle evolution^{29–31}. □

Received 30 November 1993; accepted 3 May 1994.

- Ridgen, S., Ahrens, T. J. & Stolper, E. M. *Science* **226**, 1071–1074 (1984).
- Agee, C. B. & Walker, D. J. *geophys. Res.* **93**, 3437–3449 (1988).
- Miller, G. H., Stolper, E. M. & Ahrens, T. J. *geophys. Res.* **96**, 11831–11848 (1991).
- Stolper, E., Walker, D., Hager, B. H. & Hays, J. F. *J. geophys. Res.* **86**, 6261–6271 (1981).
- Ohtani, E. *Earth planet. Sci. Lett.* **67**, 261–272 (1984).
- Miller, G. H., Stolper, E. M. & Ahrens, T. J. *geophys. Res.* **96**, 11849–11864 (1991).
- Agee, C. B. & Walker, D. *Earth planet. Sci. Lett.* **90**, 144–156 (1988).
- Jordan, T. H. *Phil. Trans. R. Soc. A* **304**, 359–373 (1981).
- Menzies, M. A. & Hawkesworth, C. J. (eds) *Mantle Metasomatism* (Academic, London, 1987).
- Eggler, D. H. & Wendlandt, R. F. in *Kimberlites, Diatremes and Diamonds: Their Geology, Petrology, and Geochemistry* (eds Boyd, F. R. & Mayer, H. O. A.) 330–338 (Am. Geophys. Un., Washington DC, 1979).
- Canil, D. & Scarfe, C. M. *J. geophys. Res.* **95**, 15805–15816 (1990).
- Williams, Q. & Jeanloz, R. *Science* **239**, 902–905 (1988).
- Stixrude, L. & Bukowski, M. S. *Science* **250**, 541–543 (1990).
- Revenaugh, J. & Jordan, T. H. *J. geophys. Res.* **94**, 5787–5813 (1989).
- Revenaugh, J. & Jordan, T. H. *J. geophys. Res.* **96**, 19749–19762 (1991).
- Revenaugh, J. & Jordan, T. H. *J. geophys. Res.* **96**, 19763–19780 (1991).
- Revenaugh, J. & Jordan, T. H. *J. geophys. Res.* **96**, 19781–19810 (1991).
- Gasparik, T. *Contr. Miner. Petrol.* **102**, 389–404 (1989).
- Sipkin, S. A. & Revenaugh, J. *J. geophys. Res.* (in press).
- Anderson, D. L. *Theory of the Earth* (Blackwell Scientific, Boston, Massachusetts, 1989).
- van der Hilst, R. D., Engdahl, E. R. & Spakman, W. *Geophys. J. Int.* **115**, 264–302 (1993).
- Dziewonski, A. M., Hales, A. L. & Lapwood, E. R. *Phys. Earth planet. Inter.* **10**, 12–48 (1975).
- O'Connell, R. J. & Budiansky, B. *J. geophys. Res.* **82**, 5719–5735 (1977).
- Mavko, G. M. *J. geophys. Res.* **85**, 5173–5189 (1980).
- Toramaru, A. & Fujii, N. *J. geophys. Res.* **91**, 9239–9252 (1986).
- Spiegelman, M. & McKenzie, D. *Earth planet. Sci. Lett.* **83**, 137–152 (1987).
- Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **25**, 297–356 (1981).
- Ohtani, E. *Earth planet. Sci. Lett.* **67**, 261–272 (1984).
- O'Nions, R. K., Evensen, N. M. & Hamilton, P. J. *J. geophys. Res.* **84**, 6091–6101 (1979).
- Jacobsen, S. B. & Wasserburg, G. J. *J. geophys. Res.* **84**, 7411–7427 (1979).
- DePaolo, D. J. *Geochim. cosmochim. Acta* **44**, 1185–1196 (1980).
- McKenzie, D. *Earth planet. Sci. Lett.* **95**, 53–72 (1989).

ACKNOWLEDGEMENTS. We thank Q. Williams, J. Vidale, P. Shearer and J. Phipps Morgan for comments and criticism. This research was supported by the US National Science Foundation.

Burgess Shale-type fossils from a Lower Cambrian shallow-shelf sequence in northwestern Canada

Nicholas J. Butterfield

Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

Fossil Lagerstätten comparable to that of the Burgess Shale potentially provide the detail necessary to resolve and assess the so-called Cambrian Explosion of multicellular life^{1,2}; distributional and taphonomic biases, however, often limit the generalizations that can be drawn. Here I report widespread occurrences of Burgess Shale-type fossils in shallow-shelf sediments of the Lower Cambrian Mount Cap Formation, District of Mackenzie, Northwest Territories, Canada. These borehole assemblages significantly expand the known geographical and ecological range of such fossil biotas and document the early appearance of both wiwaxiid polychaetes and filter-feeding crustaceans. The ability of this latter group to exploit microplanktic primary productivity marks a fundamental shift in trophic structuring and may distinguish Phanerozoic from pre-Phanerozoic ecosystems.

The Mount Cap succession typically comprises 100–300 m of carbonates and conspicuously glauconitic sandstones, siltstones and shales³, an intermediate- to shallow-shelf setting previously undersampled at this taphonomic level⁴. In the subsurface of the Coleville Hills area it lies 200–300 km landward of the shelf margin, and entirely outside the zone of Cordilleran deformation (Fig. 1); it is essentially flat-lying and unaltered, with an alginite horizon thought to be a potential, though immature, petroleum source rock⁵. Organic-walled fossils are exceptionally well preserved, with structures as narrow as 0.1 µm fully resolvable; they are typically light orange in colour and variably fluorescent⁵. By contrast, the Burgess Shale and Sirius Passet (Greenland)⁶ biotas have been exposed to significant metamorphic heating, whereas the Chengjiang fossils of China are severely weathered.

The Mount Cap Formation straddles the Lower/Middle Cambrian boundary; however, all the fossils discussed here come

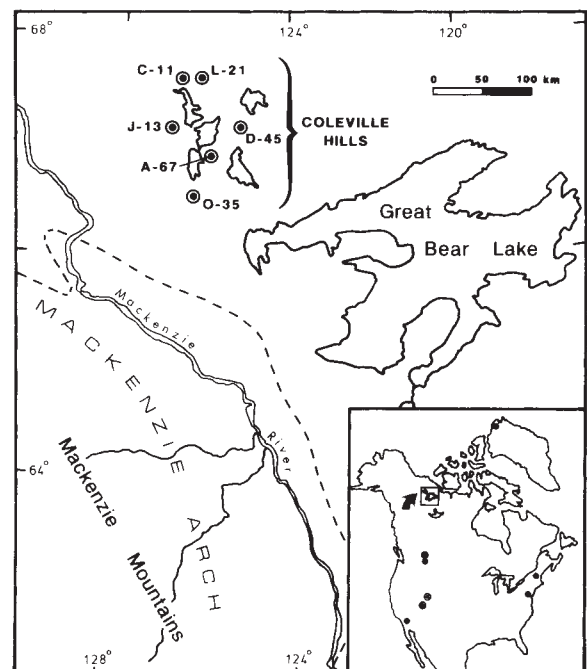
from the lowermost parts of the sequence, falling within the Lower Cambrian *Bonnina-Olenellus* trilobite zone³. The presence of the trilobite *Wanneria* 20 m above the most productive sample (Bele O-35 at 1,351 m) further constrains the stratigraphic position to at least the middle of that zone^{3,7}, placing this material among the oldest Burgess-type biotas known from Laurentia⁴. Even so, the Mount Cap assemblage appears to belong to a broadly continuous belt of comparable Lower and Middle Cambrian fossil biotas extending from southern California⁸ through to northern Greenland⁶ and Pennsylvania⁹ (Fig. 1).

Green and black shales and siltstones of Mount Cap drill-core and drill-cuttings were examined for organic-walled (Burgess Shale-type) fossils, both on bedding surfaces, and using a low-manipulation hydrofluoric acid maceration technique¹⁰ (with bedding-parallel thin-sections to confirm authenticity). Nine significantly fossiliferous samples from six boreholes have so far been identified, separated laterally by up to 125 km (Fig. 1) and stratigraphically by at least 25 m. Macroscopic compressions include a palaeoscolecidan worm, a partial *Anomalocaris* appendage, and several ~2-cm-diameter circular structures with pronounced radial and concentric fabrics, possibly scenellids. Finely preserved shelly fossils include trilobites and lingulids with extractable organic layers, and articulated(?) *Protospongia* spicules.

It is the smaller fossils, however, that provide the more interesting data. Chaetae (paleae) of *Wiwaxia*, a jawed polychaete¹¹ previously known only from the Middle Cambrian Stephen Formation (including the Burgess Shale) and the similarly aged Spence Shale of Utah^{12,13}, occur throughout the Coleville Hills area (in five samples from four boreholes; Figs 1 and 2a, b). Some of these are conspecific with the Burgess Shale form *W. corrugata*; however, four samples (from four boreholes) yield specimens with pronounced transverse rows of tubercles developed on the paleae blades (Fig. 2b), arguably a new species. Two *Wiwaxia*-producing samples also yielded distinctive papillose spines similar to those co-occurring with *Wiwaxia* paleae in the Burgess Shale, and may represent the fossil polychaete's 'missing' neurochaetae^{11,14}.

Small, largely disarticulated arthropod fossils occur in four samples (from three boreholes) of the Mount Cap, abundantly so in a 5-cm band of black shale/siltstone near the base of the formation (Bele O-35 at 1,351 m). Flattened cuticular remains include rare jointed appendages (Fig. 2c, l) and articulation sur-

FIG. 1 Location of boreholes intercepting fossiliferous Mount Cap Formation. The dashed line marks the easternmost extent of major Mesozoic deformation; the Mackenzie Arch is broadly coincident with the Cambrian shelf edge³. Fossiliferous samples: O-35-1331 (O-35 at 1,331 m depth below the surface), *Wanneria* cuticle; O-35-1337, *Anomalocaris*; O-35-1351, probable branchiopod crustaceans, *Wiwaxia*, O-35-1352, *Wiwaxia* and associated papillose spines, palaeoscolecidan worm; C-11-1235, arthropods, 'algal' filaments; C-11-1260, *Wiwaxia*, arthropods, 'algal' filaments; J-13-1240, arthropods, 'algal' filaments; D-45-937, *Wiwaxia* and associated papillose spines; L-21-1124, *Wiwaxia*; A-47-1287, circular structures (scenellids?). The reference map includes the principal Laurentian Burgess Shale-type fossil localities, with filled symbols indicating Lower Cambrian occurrences and semi-filled symbols Middle Cambrian occurrences. Those of early Cambrian age all appear to belong to the upper Lower Cambrian *Bonnina-Olenellus* zone⁴. *Wanneria*, a mid-zone trilobite^{3,7}, occurs in the lower Mount Cap shales and the Kinzers (Pennsylvania)⁹ and Eager (Cranbrook, BC) formations, but occurs below fossiliferous units in the Latham Shale (southern California)⁸. The age of the Sirius Passet biota in the upper Buen Formation (north Greenland)⁶ is less well resolved: Buen trilobites fail to distinguish between the upper Lower (*Bonnina-Olenellus* zone) or middle Lower (*Nevadella* zone) Cambrian²⁹; *Wanneria*, however, occurs in the lowermost unit of the conformably overlying Brøndum Fjord Group²⁹, and shales considered to lie stratigraphically below the Sirius Passet biota contain acritarchs which, on the East European Platform, do not occur below the upper Lower Cambrian (Rausve = *Protolenus* = *Bonnina-Olenellus* zone)^{28,30}.



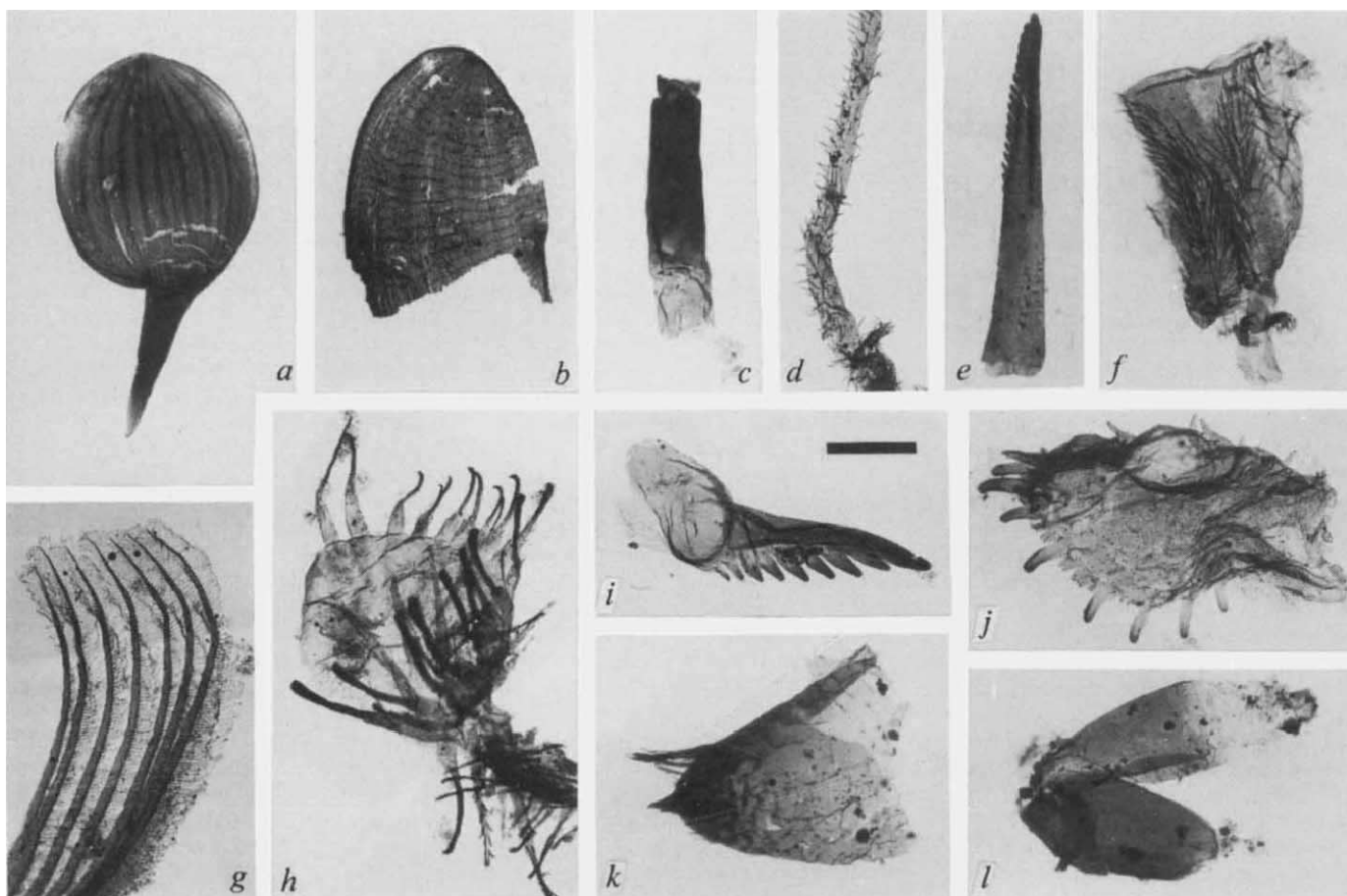


FIG. 2 Light micrographs of isolated Mount Cap *Wiwaxia* and arthropods. Figured specimens stored in the Type Collection of the Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario, Canada. Note that all specimens are flattened and most include euhedral pyrite crystals, thus confirming their syngenicity with the host sediment; comparable structures are also observed in petrographic thin-section. *a*, *Wiwaxia corrugata* palea (GSC 108957/C-248325 (Bele O-35 drill-core at 1,351 m) (GSC figured specimen no./GSC locality no.)). *b*, Partial *Wiwaxia* palea with prominent transverse rows of tubercles (GSC 108958/C-248326 (North Colville L-21 drill-core at 1,124 m)). *c*, Arthropod appendage with intact articulation (GSC 108959/C-248325). *d*, Spinose seta (108960 / C-248325). *e*, Biseriately dentate claw or spine (108961/C-248325). *f*, Cuticular fragment with setae oriented in a 'one-way' probable food-channel (108962/C-248325). *g*, Partially intact

filter plate of bipectinate setae (compare with Fig. 3a, b) (108963/C-248325). *h*, Spinose, setose and setulose appendage comparable to the trunk limbs of *Eurycercus* and other modern branchiopods (compare with Fig. 25 of ref. 16) (108964/C-248325). *i*, Uniseriately dentate spine, with articulation surface (108965/C-248325). *j*, Dentate appendage with apparent articulation surface (108966/C-248325). *k*, Terminal structure with apical spine, hair-like setae and fine combs of sub-micrometre setulae (108967/C-248327 (Ewekka C-11 drill-cuttings at 1,235 m)). *l*, Articulated podomeres with punctate and 'terraced' microstructure (108968/C-248328 (Ewekka C-11 drill-cuttings at 1,260 m)). Scale bar shown in *i* represents: 210 μ m in *b*; 120 μ m in *a*; 80 μ m in *d*, *e*, *f*; 65 μ m in *h*; 50 μ m in *g*, *i*, *j*, *l*; and 40 μ m in *c*, *k*.

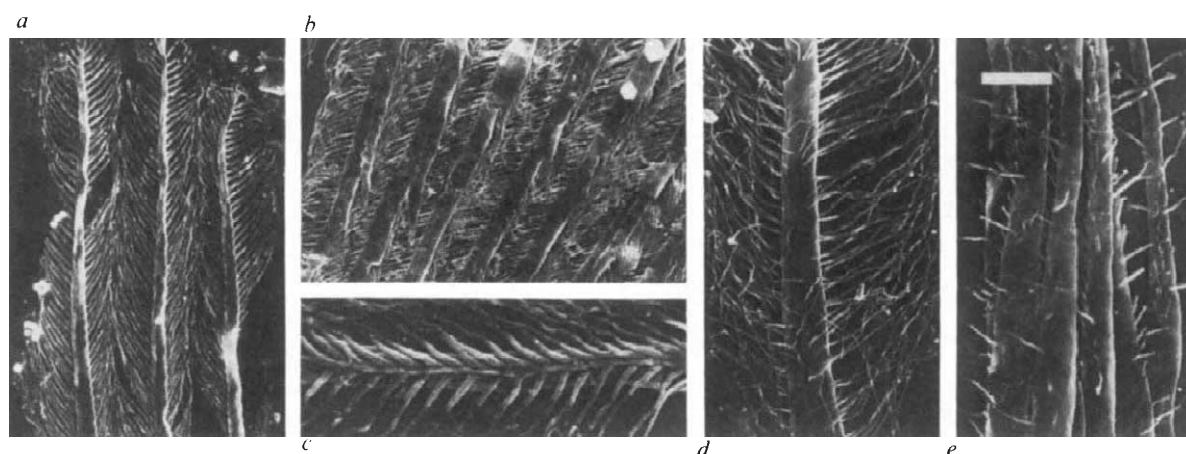


FIG. 3 Scanning electron micrographs of Mount Cap crustacean setae. Figured specimens stored in the Type Collection of the Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario, Canada. Note that all specimens are flattened and most include euhedral pyrite crystals, thus confirming their syngenicity with the host sediment; comparable structures are also observed in petrographic thin-section. *a*, Partial filter plate of regularly bipectinate setae, 'posterior' surface; setule interval $\sim 1 \mu$ m (compare with Figs 100 and 101 of ref. 19) (GSC 108969/C-248325 (Bele O-35 drill-core at 1,351 m)). *b*, Filter plate of regularly bipectinate

setae; note the serrated spurs developed on the 'posterior' margins (compare with the brush spinules of modern *Eurycercus*; ref. 16, Figs 29–31); setule interval $\sim 1 \mu$ m (GSC 108970/C-248325). *c*, 'Anterior' surface of regularly bipectinate seta; setule interval $\sim 1 \mu$ m (GSC 108971/C-248325). *d*, Plumose seta with randomly oriented hair-like setulae 0.1–0.4 μ m diameter (GSC 108972/C-248325). *e*, Group of setae with heteromorphic setulae (GSC 108973/C-248325). Scale bar shown in *e* represents 10 μ m in *a*, *b*, *d* and *e*, and 3 μ m in *c*.

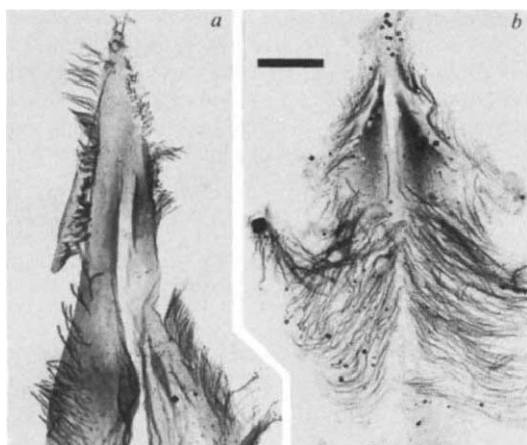


FIG. 4 Light micrographs of Mount Cap fossil arthropod labrums with setulate margins, a feature diagnostic for the Crustacea *sensu stricto*²¹. Note the presence of euhedral pyrite crystals and fossil flattening. *a*, GSC 109200/C-248325 (Bele O-35 drill-core at 1,351 m). *b*, GSC 109201/C-248325. Scale bar, 50 μ m.

faces (Fig. 2*i, j*), and a wide array of structures ornamented with hairs, tubercles, spines and/or setae (Figs 2–4); setulose setae are particularly common and include diverse plumose, denticulate and spinose forms. Intact series of up to 40 interlinked and regularly bi-pectinate setae form precise micrometre-scale filter plates (Figs 2*g* and 3*a–c*) which, in association with other specialized constituents (for example, Figs 2*f, h* and 3*b*), show a precise comparison with the filter-feeding apparatus of modern branchiopod crustaceans (G. Fryer, personal communication, and refs 15–20); the presence of a labrum with setulate margins (Fig. 4) establishes this material as crustacean at the very least²¹. As probable branchiopods, these Mount Cap fossils add significant new data to the analysis of Cambrian arthropod phylogeny, particularly as small branchiopod crustaceans are well documented from the Upper Cambrian *orsten* of Sweden²²; this close relationship with the *orsten* forms also provides an important link between fossil *Lagerstätten* of Burgess Shale-type and later taphonomic modes²³.

Setose cuticle fragments and individual setae of the Mount Cap crustaceans measure up to 3 mm in length, indicative of whole animals perhaps as long as 10 mm. However, like the *orsten* material, a wide range in size and form points to the presence of several ontogenetic stages; intact filter plates, for example, vary in size by a factor of three (180 $\times$ 130 μ m to 585 $\times$ 442 μ m). Given the precise, micrometre-scale nature of the filters, it is probable that these branchiopod-like crustaceans were clear-water filter-feeding nekton/plankton for a significant portion of their life cycle. By contrast, the arthropodan material from the C-11 and J-13 boreholes does not include filtering structures (Figs 1 and 2*k, l*), and the articulated podomeres (Fig. 2*l*) are more suggestive of a benthic, ambulatory mode of life.

Despite their morphological diversity and broad ecological partitioning, Cambrian metazoans have conventionally been interpreted as trophic generalists^{24,25} with putative filter-feeding nekton²⁶ being relatively non-specialized. Identification of an elaborate and essentially modern crustacean filter apparatus by the early Cambrian points not only to the remarkably advanced state of the Crustacea at that time^{22,27}, but also the probability that such forms constituted an essential link in the trophic webs of the earliest Palaeozoic, just as they do today. Indeed, the early exploitation of oceanic micro-, nanno- and pico-plankton by such filter-feeders would have fundamentally altered global trophic structures, conceivably fuelling the 'explosion' of Cambrian metazoan evolution. In this light the coincident radiation of planktic protists, most notably the acanthomorphic acritarchs²⁸, becomes explicable as an adaptive response to micro-grazing activity. □

Received 22 February; accepted 25 April 1994.

- Conway Morris, S. *Science* **246**, 339–346 (1989).
- Gould, S. J. *Wonderful Life* (Norton, New York, 1989).
- Aitken, J. D., MacQueen, R. W. & Usher, J. L. *Geol. Surv. Can. Pap.* **73-9**, 1–163 (1973).
- Conway Morris, S. *Trans. R. Soc. Edinb., Earth Sci.* **80**, 271–283 (1989).
- Wielens, J. B. W., von der Dick, H., Fowler, M. G., Brooks, P. W. & Monnier, F. *Bull. Can. Petrol. Geol.* **38**, 236–245 (1990).
- Conway Morris, S., Peel, J. S., Higgins, A. K., Soper, N. J. & Davis, N. C. *Nature* **326**, 181–183 (1987).
- Palmer, A. R. & Repina, L. N. *Univ. Kansas Paleontol. Contrib. (n. ser.)* **3**, 1–35 (1993).
- Briggs, D. E. G. & Mount, J. D. *J. Paleont.* **56**, 1112–1118 (1982).
- Resser, C. E. & Howell, B. F. *Bull. geol. Soc. Am.* **49**, 195–248 (1938).
- Butterfield, N. J. *Paleobiology* **16**, 272–286 (1990).
- Butterfield, N. J. *Paleobiology* **16**, 287–303 (1990).
- Conway Morris, S. *Phil. Trans. R. Soc. Lond.* **B307**, 507–586 (1985).
- Conway Morris, S. & Robison, R. A. *Univ. Kansas Paleont. Contrib. Pap.* **122**, 1–48 (1988).
- Conway Morris, S. *J. geol. Soc. Lond.* **149**, 631–636 (1992).
- Cannon, H. G. *Phil. Trans. R. Soc. Lond.* **B222**, 267–352 (1933).
- Fryer, G. *Trans. R. Soc. Edinb.* **65**, 335–381 (1963).
- Fryer, G. *Phil. Trans. R. Soc. Lond.* **B269**, 137–274 (1974).
- Fryer, G. *Phil. Trans. R. Soc. Lond.* **B303**, 229–343 (1983).
- Fryer, G. *Phil. Trans. R. Soc. Lond.* **B331**, 1–99 (1991).
- Watts, E. & Petri, M. *J. nat. Hist.* **15**, 463–473 (1981).
- Walossek, D. & Müller, K. J. *Lethaia* **23**, 409–427 (1990).
- Walossek, D. *Fossils and Strata* **32**, 1–202 (1993).
- Briggs, D. E. G. *Acta zoologica* **73**, 293–300 (1992).
- Sepkoski, J. J. & Sheehan, P. M. in *Biotic Interactions in Recent and Fossil Benthic Communities* (eds Tevesz, M. J. S. & McCall, P. L.) 673–717 (Plenum, New York, 1983).
- Conway Morris, S. *Palaeontology* **29**, 423–467 (1986).
- Briggs, D. E. G. & Whittington, H. B. *Trans. R. Soc. Edinb., Earth Sci.* **76**, 149–160 (1985).
- Sanders, H. L. in *Phylogeny and Evolution of Crustacea* (eds Whittington, H. B. & Rolfe, W. D. I.) 163–175 (Mus. comp. Zool. spec. Publ., Cambridge, 1963).
- Moczyłowska, M. *Fossils and Strata* **29**, 1–127 (1991).
- Palmer, A. R. & Peel, J. S. *Bull. Grøn. geol. Unders.* **164**, 1–35 (1993).
- Vidal, G. & Peel, J. S. *Rapp. Grøn. geol. Unders.* **91**, 29–36 (1979).

ACKNOWLEDGEMENTS. I thank D. Cook and the Institute of Sedimentary and Petroleum Geology, Calgary, for access to borehole materials and a contribution to laboratory costs, and G. Budd, G. Fryer, S. Conway Morris, S. Bengtson, D. Briggs and H. Whittington for discussion. This research is supported by a Trevelyan Research Fellowship, Selwyn College, Cambridge.

Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex

Adam M. Sillito, Helen E. Jones,
George L. Gerstein* & David C. West

Department of Visual Science, Institute of Ophthalmology,
Bath Street, London EC1V 9EL, UK

* Department of Neuroscience, School of Medicine,
University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

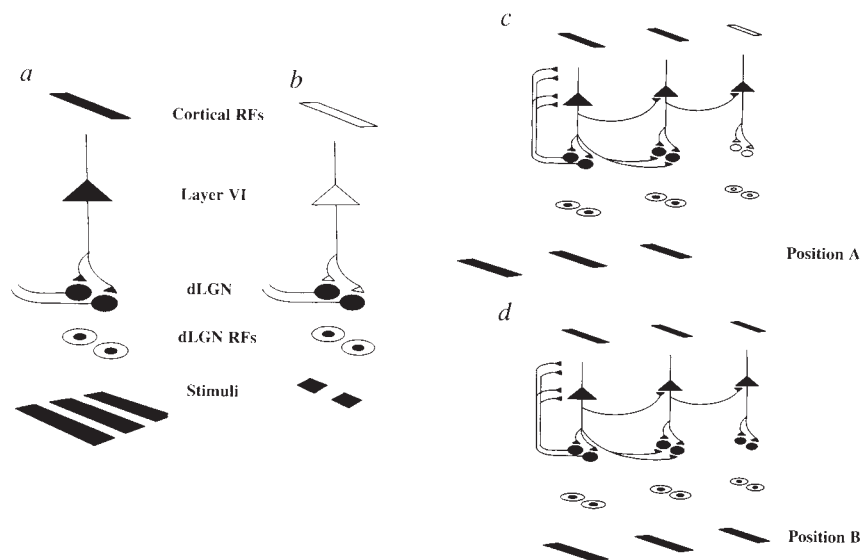
THE function of the massive feedback projection from visual cortex to its thalamic relay nucleus^{1,2} has so far eluded any clear overview. This feedback exerts a range of effects^{3–6}, including an increase in the inhibition elicited by moving contours^{7,8}, but the functional logic of the direct connections to the thalamic cells that relay the retinal input to the cortex^{9–11} remains largely unknown. In contrast to its thalamic nucleus, the visual cortex is characterized by cells that are strongly sensitive to the orientation of moving contours. Here we report that when driven by moving oriented visual stimuli the cortical feedback induces correlated firing in relay cells. This cortically induced correlation of relay cell activity produces coherent firing in those groups of relay cells with receptive field alignments appropriate to signalling the particular orientation of the moving contour to the cortex. Synchronization of relay cell firing means that they will elicit temporally overlapping excitatory postsynaptic potentials in their cortical target cells, thus increasing the chance that the cortical cells will fire. Effectively this increases the gain of the input for feature-linked events detected by the cortex. We propose that this feedback loop serves to lock or focus the appropriate circuitry onto the stimulus feature.

Figure 1 illustrates the essence of our approach and shows feedback connections from a cell in layer VI of the visual cortex

to relay cells in the dorsal lateral geniculate nucleus (dLGN), the thalamic nucleus relaying the retinal input to the cortex. The key prediction is that when the cortical feedback cells tuned to any particular orientation discharge, they will tend to cause the dLGN cells they contact to fire together because they provide a common input. We suggest, however, that the cortico-thalamic

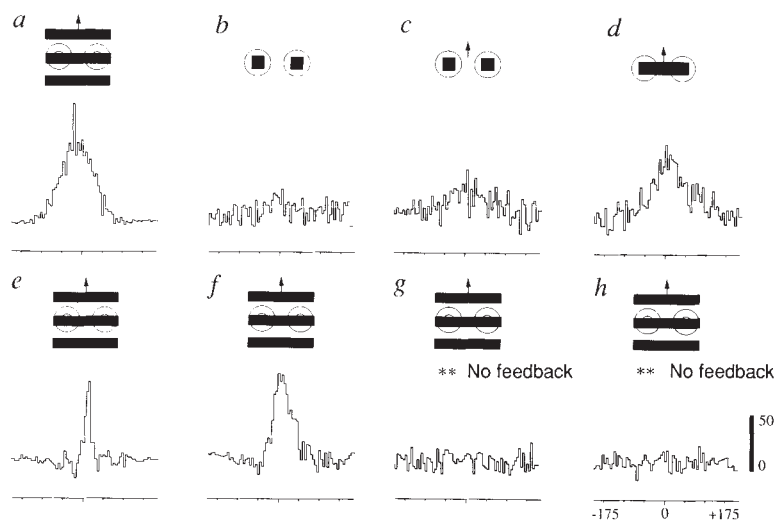
input is only strong enough to exert an effect on those dLGN cells that are additionally depolarized by their retinal input. Logically these will be the dLGN cells synchronously activated by the contour driving the cortical cell. Although we indicate specific connections from layer-VI cells to pairs of appropriately aligned dLGN cells, this is only to illustrate the concept of the

FIG. 1 The broad logic of the experiment is illustrated by the schematic diagrams. The cortical receptive fields (cortical RFs) refer to the layer-VI cells providing feedback to the dorsal lateral geniculate nucleus (dLGN); the receptive fields of the dLGN cells (dLGN RFs) are shown underneath and driven either by the drifting grating (a) or stationary flashing squares (b). The shaded areas in the schematic indicate the elements of the system driven by the two stimuli. The work was carried out on the feline visual system and details of anaesthesia and preparation are given elsewhere⁸. We recorded from groups of three or more dLGN cells with tripartite metal electrode assemblies, constructed so that for eccentricities within 13° of the area centralis they picked up receptive field locations separated by between 1° and 4° . Data were then taken for the responses to a drifting sinusoidal grating at orientations selected to align respectively with each of the pairs of fields. We constructed cross correlograms to examine temporal links in the firing of the cell pairs. Schematics c and d illustrate excitatory connections additional to those shown in a and b; these will influence the correlations in Figs 2 and 3. The filled profiles represent the elements activated by the stimulus. At stimulus position 'A', the collaterals between like orientation columns introduce a coordinating influence and will affect adjacent locations in the path of the stimulus. Thus there will be an anticipatory coordinating influence on the cells about to be activated at position 'B'. The connections shown here are purely to illustrate the concept of the interaction not the actual circuitry. Although we indicate specific connections from layer-VI cells to pairs of appropriately aligned dLGN cells, this is only to illustrate the concept of the interaction. The projection from a layer-VI cell could be diffusely distributed over a patch of dLGN cells. The stimulus selects the specificity by the activation of layer-VI cells tuned to its orientation and the timing of the depolarization it produces in dLGN cells. The option for synchronization comes when two dLGN cells are



simultaneously depolarized and receive common input from layer-VI cells driven by the stimulus. The timing of the loop from dLGN to cortex and back allows for cortical feedback within 5–10 ms of the onset of the response, enabling the feedback coordination to contribute to the development of the effective strength of the input to the cortex. When the stimulus is moving thus, both the anticipatory effect from connections between adjacent columns and the in-line loop will exert an influence on the synchronization of the discharge in relay cells. Inhibitory interneurons receive a significant component of the corticofugal input and, although not shown here, will also play a part in the effects seen. A common mode inhibitory input to two cells will cause a weak but broad peak in the cross correlogram. The effects reported here will reflect the interplay of all these factors.

FIG. 2 Examples of cross correlograms from pairs of dLGN cells with and without corticofugal feedback. These are corrected for direct stimulus coordination by sequentially shifting in time the spike train elicited from a trial for one of the cells by one trial through to the total number of trials, and averaging the cross correlation produced for each of these shifts. This is then subtracted from the raw cross correlogram^{18–21}. Examples of the data obtained are shown in records a–g. The timescale refers to ± 175 ms either side of the zero point. The vertical calibration to the right of record f shows the range corresponding to 0–50 counts per bin. Records a and b show the cross correlograms obtained from a pair of on-centre X cells driven respectively by a drifting grating and two stationary flashing squares. The grating would drive layer-VI cells well but the flashing squares would not. e and f. Further examples of the correlations obtained from the use of a grating for two off-centre X cells (e) and two on-centre Y cells (f). Removal of the feedback by aspiration of areas 17, 18 and 19 (ref. 3), although not diminishing the excitatory response to a grating stimulus, eliminated the neural coordination in the cross correlogram, as shown in g and h, respectively, for pairs of off-centre Y cells and on-centre X cells. c and d. The same pair of on-centre X cells as in a and b, but showing the correlations obtained when the stimuli were drifting squares (c) or a drifting bar (d). In total we examined 53 cell pairs. Of 37 cell pairs studied in the presence of feedback, 19 were between identical centre polarity X and Y cell pairs and 15 showed



significant correlations. Of the remaining 18 mixed pairs, 4 showed correlations. In the absence of feedback we examined 16 cell pairs, of which 10 were identical types and no pair showed significant correlations.

interaction. The projection from a layer-VI cell could be diffusely distributed over a patch of dLGN cells. The stimulus selects the specificity by the activation of layer-VI cells tuned to its orientation and the timing of the depolarization it produces in dLGN cells. The option for synchronization comes when two dLGN cells are simultaneously depolarized and receive common input from layer VI cells driven by the stimulus.

We recorded from pairs of dLGN cells and looked for correlated firing when they were stimulated by a drifting grating aligned so as to stimulate their two fields synchronously. The degree of correlation was checked by constructing cross-correlograms of their firing patterns as shown in Fig. 2*a, e* and *f*. They all show a strong statistically significant peak indicating correlated firing. In this type of test it is necessary to distinguish the simple coordinated firing of the dLGN cells caused by the synchronous stimulation from that caused by a common feedback input (Fig. 1*a, b*). The cross correlograms shown in Fig. 2 are thus corrected for the stimulus-induced effects (see Fig. 2 legend for details). We checked the possibility that the common input derived from subcortical connections by performing two tests. First, we examined the responses to stimuli known to drive dLGN cells effectively but not layer-VI cortical cells; synchron-

ous flashing of small squares on the receptive field centres of the two dLGN cells caused responses of the same magnitude as the drifting grating, but did not cause a peak in the cross correlogram (Fig. 2*b*). Second, when we tested for correlations in the responses to a drifting grating with the visual cortex removed, there were none (Fig. 2*g, h*). Interestingly, in the presence of feedback, the cell pair shown in Fig. 2*a* and *b* did reveal a hint (not statistically significant) of a correlation when the two squares were drifted rather than flashed over the field (Fig. 2*c*) and these correlations became statistically significant when the drifting squares were extended so as to form a bar (Fig. 2*d*). The degree of correlation closely mirrors the responsiveness of cortical layer-VI cells to the three different stimuli.

The cross correlograms shown in Fig. 2 give limited information because they concatenate events over the entire stimulus period and so do not register variation in correlation during the stimulus period. We constructed joint peristimulus time histograms (JPSTHs) (Fig. 3) to show the variation in the significance of the correlation through the stimulus period. The key points are that the times of the spikes during the stimulus period for the two cells form the *x*- and *y*-axes of a matrix, effective stimulus onset is at the origin, and the colour of each bin (small square) in the matrix shows the significance level of the correlation at that point, with white for the highest positive correlation and black for those points where the correlation falls most below what would have been expected from chance (Fig. 3 legend). The points where the two cells fire simultaneously fall along the diagonal of the square formed by the matrix, those where the cell on the *y*-axis fires before or after the other fall below and above the diagonal, respectively. Thus for each moment during the response to the stimulus, the squares of the matrix show whether there were significant correlations for the cells firing simultaneously, or one firing at some interval before or after the other. The diagonal distribution of white squares in Fig. 3*A* shows a strong focus of synchronous and near-synchronous firing, whereas the banding of black squares indicates that certain intervals in the firing between the two cells have a less than chance representation. The coincidence histogram to the right of the matrix summarizes the way the coincident and near-coincident points vary during the stimulus cycle. A more complex

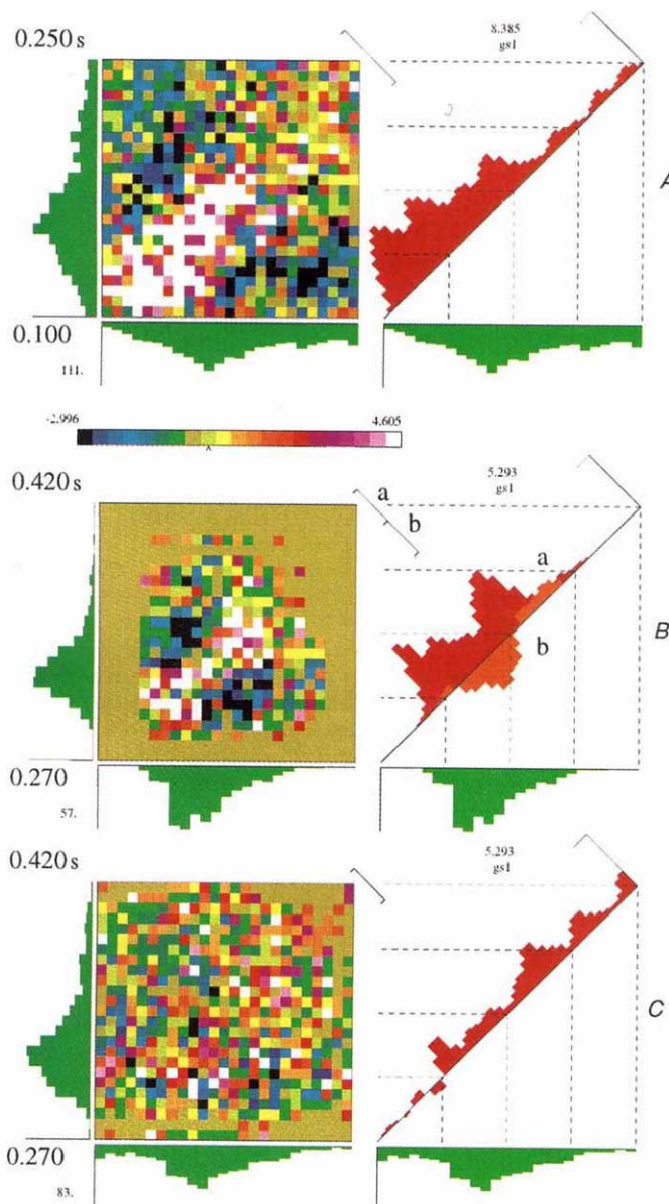


FIG. 3 These three joint PSTHs show the variation in correlation significance through the stimulus period for two pairs of dLGN cells with corticofugal feedback (A, B) and one pair without (C). Positive correlations of the highest significance are shown by white squares and negative correlations by black squares. The colour value under A shows the colours used to represent intermediate values of significance. A, data for two on-centre Y cells; B, off-centre X cells; C, off-centre Y cells without feedback. In the JPSTH, the two spike trains relating to the stimulus period are laid out along the *x* and *y* axis, with time of each successive stimulus onset at the origin. In the matrix, the points corresponding to the logical 'AND' are accumulated and displayed at an appropriate resolution in terms of bin size. Points along the diagonal represent the time-locked average for near coincidences. These form the peristimulus coincidence histogram and are shown by the diagonal to the right of the matrix. The plot subsumes two levels of development: (1) a correction for direct stimulus modulation of the two firing rates by subtracting the bin-by-bin product of the two PSTHs and then dividing by the bin-by-bin product of the standard deviations of the two PSTHs (this normalizes the JPSTH); (2) a significance test based on a bin-by-bin calculation of the possible count distributions in a JPSTH bin, noting the actual count in the bin, the counts in corresponding bins of the PSTHs and the number of stimulus repetitions. The display is generated as the measure for high counts minus that for low counts, thus establishing the significantly high and low count points (bins). The measure used is the negative natural logarithm of the probability of finding the count, with 4.605 corresponding to the 99% confidence limit for positive correlations, and -2.996 the 95% confidence limit for negative correlations. This method is described in detail elsewhere^{20,21}.

temporal dependence is shown in Fig. 3B, in which positive correlations are strongest during the rising and falling phases of the response, and the black squares form two temporal foci where the correlations fall significantly below chance level. The coincidence histograms taken for the average of those points with synchronous or near-synchronous firing (a) and the average of those points where the cell on the y axis fires before the other (b) to the right of the matrix in Fig. 3B underline the difference in the variation of the negative and positive correlations through the stimulus period. In the absence of feedback, the pattern of correlation was random (Fig. 3C). In no case did we see significant correlations between pairs of cells studied (16/16) in the absence of feedback. These results indicate that (1) the feedback effects are statistically significant, and (2) that the feedback introduces a temporal structure to the correlations that varies with the stimulus cycle and involves both increases and decreases in the probability of correlation.

Interestingly, from a sample of 37 cell pairs studied at length with feedback, the majority (80%) of those showing correlations in the presence of feedback (15/19) comprised cells of the same type, with respect to both X/Y and on/off-centre properties (thus one pair might comprise two on-centre X or two off-centre Y cells). Conversely, the majority of those showing no correlation were mixed pairs (14/18). This suggests that there may be specific feedback to the X and Y systems. Clearly the cross-correlation functions observed are more complex than would be predicted by the simplified circuitry in Fig. 1a, but this omits many known features of the system. A more detailed representation is shown in Fig. 1c and d, with excitatory connections between cells in cortical columns of the same orientation selectivity but representing adjacent regions of visual space¹². The drifting grating stimulus would simultaneously activate cortical cells of similar orientation preference in columns representing adjacent locations in visual space and thus precipitate a coordinating interaction between them¹²⁻¹⁵ and their feedback. Furthermore, the spatial spread of the axons of the layer-VI cells in the dLGN would tend to pool the coordinated influences from the spatially adjacent cortical locations^{1,2}. Additionally, the output from each orientation column would come from several cells and not from a single cell; it would also make contacts on inhibitory interneurons as well as direct contacts on relay cells. As the stimulus moves it will thus bring feedback interaction from the cortical cells representing the same location in visual space (Fig. 1 legend) and adjacent locations. We suggest our data reflect the combined neural influence of all these elements, together with the associated inhibitory circuits¹⁶. Our findings are consistent with the evidence regarding stimulus binding in the visual cortex¹³⁻¹⁵ and suggest that a component of the mechanism underlying it may reflect cortico-thalamic feedback.

The effect of the cortically induced correlated firing in dLGN cells relaying information about particular features to the visual cortex will be to increase their effectiveness in driving the cortex, because of the synchronous excitatory postsynaptic potentials (e.p.s.ps) elicited in their target cortical cells. In the context of Hubel and Wiesel's model¹⁷ for establishing orientation selectivity in layer IV, it would provide a 'looped' influence enhancing the effect of the convergent geniculate inputs in layer IV. The cortical feedback may thus 'test' for the presence of the relevant pattern in the input cells; when present and correlated, this will in turn lead to a closer temporal coupling of both cortical and input cells in relation to the unique pattern of activity elicited by the driving feature (Fig. 1c, d legend). The increased coherence in the e.p.s.ps under these conditions will tend to amplify the effectiveness of the synaptic transfer between stages and further the resolution of the neuronal ensemble engaged by the stimulus. In this view, the feedback circuit searches for correlations that support the 'hypothesis' represented by a particular pattern of cortical activity and 'locks' the ensemble onto the stimulus. We believe this provides a strong case for the functional role of cortico-thalamic feedback in visual processing and

suggest that similar principles may apply to corticofugal projections in other sensory systems. □

Received 20 January; accepted 25 April 1994.

1. Robson, J. A. *J. comp. Neurol.* **216**, 89-103 (1983).
2. Wilson, J. R., Friedlander, M. J. & Sherman, S. M. *Proc. R. Soc.* **221**, 411-436 (1984).
3. Tsumoto, T., Creutzfeldt, O. D. & Legendy, C. R. *Brain Res.* **345**, 364 (1978).
4. Marrocco, R. T. & McClurkin, J. W. *Expl Brain Res.* **59**, 50-56 (1985).
5. Varela, F. J. & Singer, W. *Expl Brain Res.* **66**, 10-20 (1987).
6. Lindstrom, S. & Wrobel, A. *Expl Brain Res.* **79**, 313-318 (1990).
7. Murphy, P. C. & Sillito, A. M. *Nature* **329**, 727-729 (1987).
8. Sillito, A. M., Cudeiro, J. & Murphy, P. C. *Expl Brain Res.* **93**, 6-16 (1993).
9. Ahlsen, G., Grant, K. & Lindstrom, S. *Brain Res.* **234**, 454-458 (1982).
10. Weber, A. J., Kalil, R. E. & Behan, M. J. *comp. Neurol.* **289**, 156-164 (1989).
11. Montero, V. M. *Expl Brain Res.* **86**, 257-270 (1991).
12. Ts'o D. Y., Gilbert, C. D. & Wiesel, T. N. *J. Neurosci.* **6**, 1160-1170 (1986).
13. Gray, C. M., König, P., Engel, A. K. & Singer, W. *Nature* **338**, 334-337 (1989).
14. Engel, A. K., König, P., Gray, C. M. & Singer, W. *Eur. J. Neurosci.* **2**, 588-606 (1990).
15. Gray, C. M., Engel, A. K., König, P. & Singer, W. *Eur. J. Neurosci.* **2**, 607-619 (1990).
16. Sillito, A. M. in *Mechanisms of GABA Action in the Visual System* (eds Mize, R. R., Marc, R. & Sillito, A. M.) 349-384 (Elsevier, Amsterdam, 1992).
17. Hubel D. H. & Wiesel T. N. *J. Physiol.* **160**, 106-154 (1962).
18. Abeles, M. *J. Neurosci. Meth.* **5**, 317-325 (1982).
19. Abeles, M. & Gerstein, G. L. *J. Neurophysiol.* **60**, 909-924 (1988).
20. Palm, G., Aertsen, A. M. H. J. & Gerstein, G. L. *Biol. Cybern.* **59**, 1-11 (1988).
21. Aertsen, A. M. H. J. et al. *J. Neurophysiol.* **61**, 900-917 (1989).

NMDA receptors and activity-dependent tuning of the receptive fields of spinal cord neurons

Gary R. Lewin^{*†}, Edward Mckintosh & Stephen B. McMahon[‡]

Department of Physiology, United Medical and Dental Schools (St Thomas's Campus), Lambeth Palace Rd, London SE1 7EH, UK

^{*} Department of Neurobiology and Behavior, State University of New York at Stony Brook, New York 11794, USA

AFTER peripheral nerve section, sensory neurons regenerate but do not regain their original topographical position in the skin¹. Here we report that in the early stages of sciatic nerve regeneration, the cutaneous receptive fields (RFs) of dorsal horn neurons are larger than normal, reflecting the disorganized topography of the regenerated afferents. When nerve regeneration is complete, small contiguous RFs emerge, indicating a central compensation for the disrupted peripheral somatotopy. If the NMDA receptor antagonist MK801 is given during regeneration, RFs do not show this reorganization, but remain large and diffuse. We suggest that the coincident activity of afferents, newly innervating adjacent or overlapping cutaneous territory, acts through postsynaptic NMDA receptors² to strengthen the central effectiveness of these inputs at the expense of other non-adjacent and non-coincidentally activated inputs. In this way, dorsal horn neurons may attain and retain restricted RFs in the face of a spatially dispersed afferent input.

The RF properties (size, location, modality) of single dorsal horn neurons were examined up to 10 weeks after section and repair of one sciatic nerve (Fig. 1) in adult rats. Neurons on the unlesioned side had normal properties and served as controls. Their RFs were on the toes in about 2/3 of cases and on the lateral foot or ankle in the remaining 1/3 (Fig. 2); the mean RF size was $118 \pm 17 \text{ mm}^2$ (mean $\pm$ s.e.; 51 neurons, 8 animals) (Fig. 3). Whole nerve recordings taken 4-5 weeks after nerve injury indicated that sciatic afferents had reinnervated targets on the lower leg but not the foot (Fig. 1c). Dorsal horn neuron RFs

[†] Present address: Department of Neurobiochemistry, Max-Planck Institute for Psychiatry, D-82152 Martinsried, Germany.

[‡] To whom correspondence should be addressed.

were diffuse and large, often occupying most of the sciatic reinnervated skin, and, as expected, were located on the leg (Fig. 2). The mean RF size was $210 \pm 23 \text{ mm}^2$ (mean $\pm$ s.e.; 35 neurons, 4 animals) (Fig. 3). This was 78% larger than control ($P < 0.01$, $F = 2.59$ d.f. = 11;74; ANOVA) and still significantly larger than control RFs located around the ankle ($P < 0.05$, unpaired t -test, $t = 2.67$, d.f. = 51). The expanded RFs probably reflected the spatial disruption of the regrown afferents. In agreement with this,

31% (11/35) of the cells had multiple RFs (never seen in controls). Where appropriate, RF areas were calculated as the sum of individual components. These changes occurred despite the fact that RFs of individual primary afferents are smaller at earlier stages of skin reinnervation^{3,4}.

In other animals, neurons were examined after 8–10 weeks of nerve regeneration when afferents reinnervate the toes, as well as the lower leg (Fig. 1c). Unlike in the earlier stages of reinner-

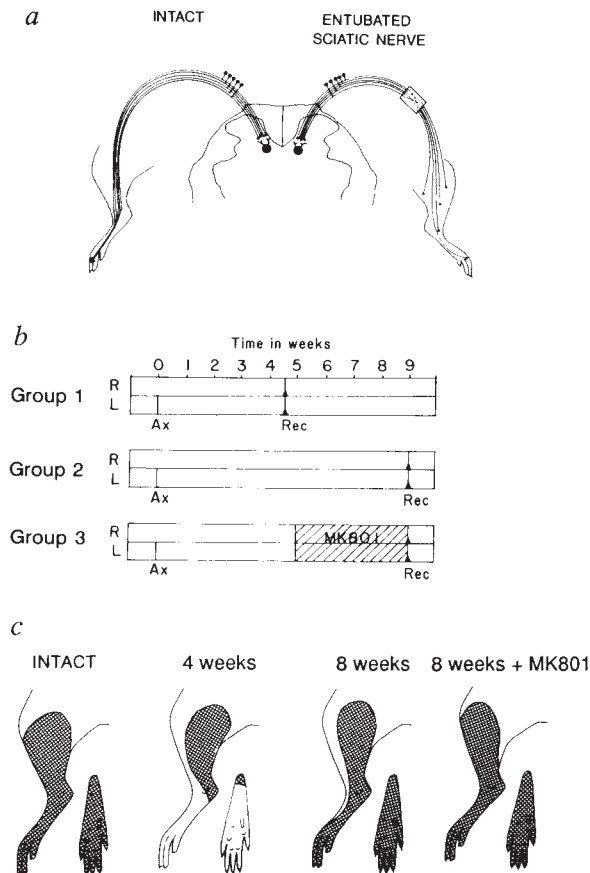


FIG. 1 Adult male rats (CSE strain) weighing 200–245 g were anaesthetized with pentobarbitone (40 mg kg^{-1} ; i.p. injected) under sterile conditions, one sciatic nerve was sectioned in the upper thigh, the cut ends were entubated, leaving a 0.5–1 mm gap to disrupt maximally the spatial organization of regenerating afferents (a). Either 4–5 or 8–10 weeks later, under urethane anaesthesia (1.25 g kg^{-1} ; i.p.), tungsten microelectrodes were used to record the activity of single dorsal horn neurons responding to hindlimb mechanical stimulation. The saphenous nerve was acutely sectioned at the beginning of the terminal experiments. The size and location of receptive fields (RFs) were drawn on scale drawings, which were later digitized to calculate RF areas. Similar experiments were performed in which MK801 was administered continuously for 4 weeks ($0.015 \text{ mg kg}^{-1} \text{ h}^{-1}$) using Alzet osmotic pumps inserted subcutaneously between the scapulae. The pump delivered $2.5 \mu\text{l}$ per h of a 0.6% solution in saline. MK801-treated animals showed an initial brief period of motor disturbances (ataxia and cumbersome locomotion) lasting 2–6 days, but the animal's behaviour recovered after 1 week. In terminal experiments on MK801-treated animals, the urethane (1.25 g kg^{-1} ; i.p.) had to be delivered slowly as MK801 appeared to potentiate its action. b, Timing of interventions in the different experimental groups. Ax, axotomy and entubation of the sciatic nerve; Rec, timing of terminal recording experiments. c, In the terminal experiments, the territory innervated by the sciatic nerve was determined by whole nerve recording and light brushing of the skin. Representative examples of these 'brush' fields from different experimental groups are shown. Note that 4 weeks after sciatic cut and repair, regenerating fibres have reached the leg but not the distal foot. By eight weeks, the entire lower leg and foot are reinnervated. The MK801 treatment does not alter the degree of reinnervation.

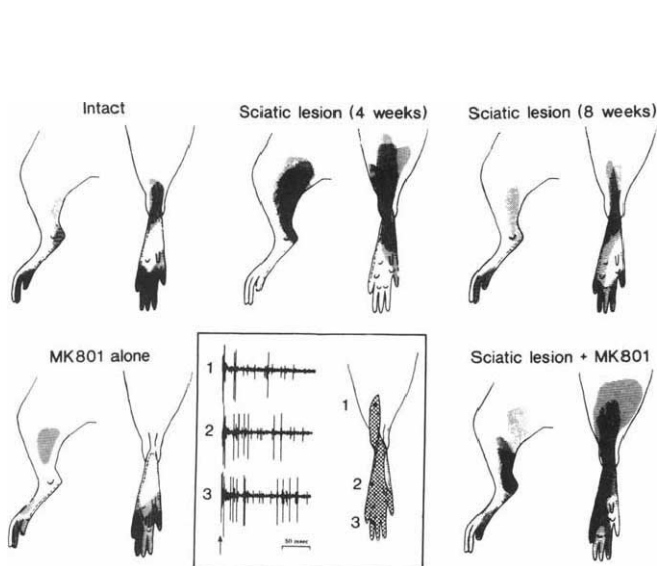


FIG. 2 Figures show the size and location of representative receptive fields of dorsal horn neurons recorded from different groups of animals. Each RF is shown by cross-hatching at a particular orientation; where RFs overlap, cross-hatching becomes more dense. Note that in intact animals the RFs are small and mostly confined to the distal foot, as expected. After sciatic nerve lesion, the RFs are initially much larger and are now found on the lower leg. With time, however, RFs become smaller, on average indistinguishable from intact animals, although not concentrated on the distal foot. MK801 treatment alone did not change RF size, whereas MK801 treatment in sciatic-lesioned animals resulted in very large RFs. Insert shows an example of one neuron recorded from an animal treated with MK801 and with a regenerated sciatic nerve. This cell had a very large receptive field determined with 'natural' stimulation (hatched area), and responded to percutaneous electrical stimulation (5 mA for 5 ms; delivered with pin electrodes) at sites proximal, intermediate and distal (labelled 1, 2 and 3) with an early (A-fibre-mediated) and late (presumed to be C-fibre-mediated) response, as shown in the traces on the left; timing of the stimulus is indicated by an arrow.

METHODS. Single dorsal horn neurons were recorded with tungsten microelectrodes (tip diameter $\sim 5 \mu\text{m}$), and depth adjustments made to optimize spike size. Signals were displayed on an oscilloscope using an analogue delay line to reveal the complete spike shape. Units were considered single on the basis of constant spike shape and amplitude, and absence of interspike intervals of less than 2 ms. RFs were mapped with a camel-hair brush, blunt glass probes and non-serrated forceps while listening to discriminated spikes on an audio amplifier.

vation, RFs were very similar to control, the mean size being $139 \pm 17 \text{ mm}^2$ (mean $\pm$ s.e., 42 neurons, 4 animals), which was indistinguishable from control ($P > 0.89$, $F = 0.51$, d.f. = 11;81; ANOVA) and significantly reduced from that seen after 4–5 weeks of regeneration ($P < 0.001$, $F = 2.43$, d.f. = 7;69; ANOVA) (Fig. 3). Only 4/42 (9.5%) of these cells had multiple RFs. Peripheral stimulation gave robust responses, and habituation (commonly seen at 4–5 weeks) was rare. Thus, it appears that spinal neurons select a new subset of afferents (those newly innervating a contiguous skin area) to form discrete RFs. The somatotopic representation of different skin areas appeared to be disrupted, as previously described^{5–7} (Fig. 2).

As functional NMDA receptors, presumptive NMDA binding sites and NMDA_{R1} messenger RNA are all present in spinal neurons^{8,9}. We postulated that the late reduction in RF size might depend on NMDA receptors. Therefore we treated animals with sciatic lesions for 4 weeks (Fig. 1b) with MK801 (ref. 10) at 0.015 mg kg^{-1} per hour. Cells on the unlesioned side of these animals were normal. The mean RF size was slightly but nonsignificantly increased to $162 \pm 17 \text{ mm}^2$ (mean $\pm$ s.e., 60 neurons, 6 animals), which was accounted for by a significant increase ($P < 0.01$, $F = 2.202$, d.f. = 26;85; ANOVA) in RF size of cells with high threshold input (Fig. 3b).

MK801 treatment produced dramatic changes on the sciatic-lesioned side. Most RFs were very large (Figs 2 and 3a), averaging $456 \pm 41 \text{ mm}^2$ (mean $\pm$ s.e., 42 neurons, 5 animals), which was significantly different from those at 8–10 weeks after nerve section without MK801, ($P < 0.0001$, $F = 8.073$, d.f. = 8;76; ANOVA). Thirty-seven per cent (15/42) of the cells had multiple RFs (compare a value of 9.5% without MK801). Thus by blocking the NMDA receptor, the dispersed spatial spread of afferent input to individual dorsal horn neurons was revealed. The MK801 treatment did not seem to alter the degree of peripheral reinnervation (Fig. 1c).

In some neurons with large RFs, we always found the responses to percutaneous electrical stimuli to be in concordance with the RFs determined by 'natural' stimulation (Fig. 2, insert).

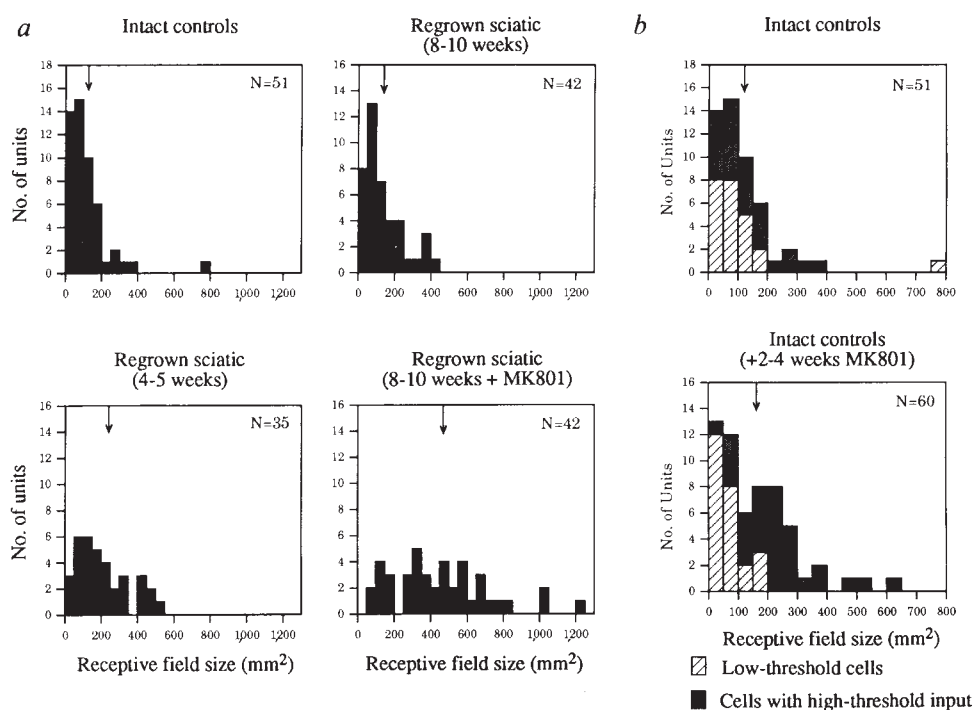
We used low doses of MK801 (compared with studies using acute doses^{11,12}), which were probably selective for NMDA receptors as RF properties contralateral to sciatic lesions were largely unaffected, whereas blockade of non-NMDA receptors reduces RF size^{13,14}. We also tested convergent neurons for wind-up (which requires NMDA receptors^{15,16}). None of 15 neurons tested after MK801 showed wind-up, compared with ~80% of control cells.

One interpretation of these results is that afferents initially randomly reinnervate skin but withdraw from topographically inappropriate targets, but other studies suggest that this is unlikely¹. We investigated this possibility using the fact that axons in the sural nerve (a branch of the sciatic) normally have a distinct anatomy, projecting predominantly through the caudal L5 dorsal root, whereas the sciatic as a whole projects more evenly through the L4 and L5 roots (Fig. 4). After sciatic nerve section, the fibres that regenerate into the sural nerve are not the original subset of sciatic afferents, but appear to be a random subset of all sciatic afferents (Fig. 4).

It is likely that MK801 acts centrally, as there is no evidence for functioning peripheral NMDA receptors^{17,18}. Indeed, fibre outgrowth during development proceeds without NMDA receptors¹⁹. Moreover, NMDA receptors are not required for normal RF expression^{13,14}, although they are required for other forms of synaptic plasticity seen in spinal neurons^{11,12,20,21}.

After random peripheral reinnervation, we envisage that individual dorsal horn neurons are initially faced with a spatially spread afferent input, resulting in large and often multiple RFs. Activity in afferents with overlapping or adjacent cutaneous RFs

FIG. 3 Receptive field sizes of dorsal horn neurons before and after sciatic lesions and with MK801 treatment. In normal animals (a, top left), neurons recorded in the L5 segment usually have small RFs, as can be seen in the RF size distribution. At early stages of sciatic nerve regeneration (a, bottom left) about half of the normal sciatic territory has been reinnervated, but despite this relatively small innervation territory RFs sizes are on average about twice as large as normal, and the RF size distribution accordingly shows a rightward skew. After reinnervation of the leg is complete (8–10 weeks), the RF distribution returns to normal (a, top right). We postulated that this reduction in RF size was due to neurons 'selecting' a subset of the spatially dispersed afferents (just those afferents that innervated adjacent territory). After afferent regeneration combined with MK801 treatment, the result was very different (a, bottom right). The RF size distribution becomes very spread, with cells having much larger RFs than normal. Thus there appears to be no compensation for the spatially dispersed input with NMDA receptor block. In b, the effects of MK801 treatment alone are shown. The overall distribution of RFs was not significantly different from untreated, intact animals. However, when cells were divided into those with and without high-threshold input, it was clear that there was a specific effect of MK801 treatment only on cells with high-threshold input. These cells showed a significant increase in RF



size. In each histogram, the arrow indicates the mean RF size of that population. The average and range of depths of recordings (in μm) for each group were: intact, 352, 50–670; 4–5-week lesion, 357, 110–680; 8–10-week lesion, 385, 72–702; and 8–10-week lesion + MK801 groups, 374, 60–725. There was no significant difference in the average depth of recording between any of the groups.

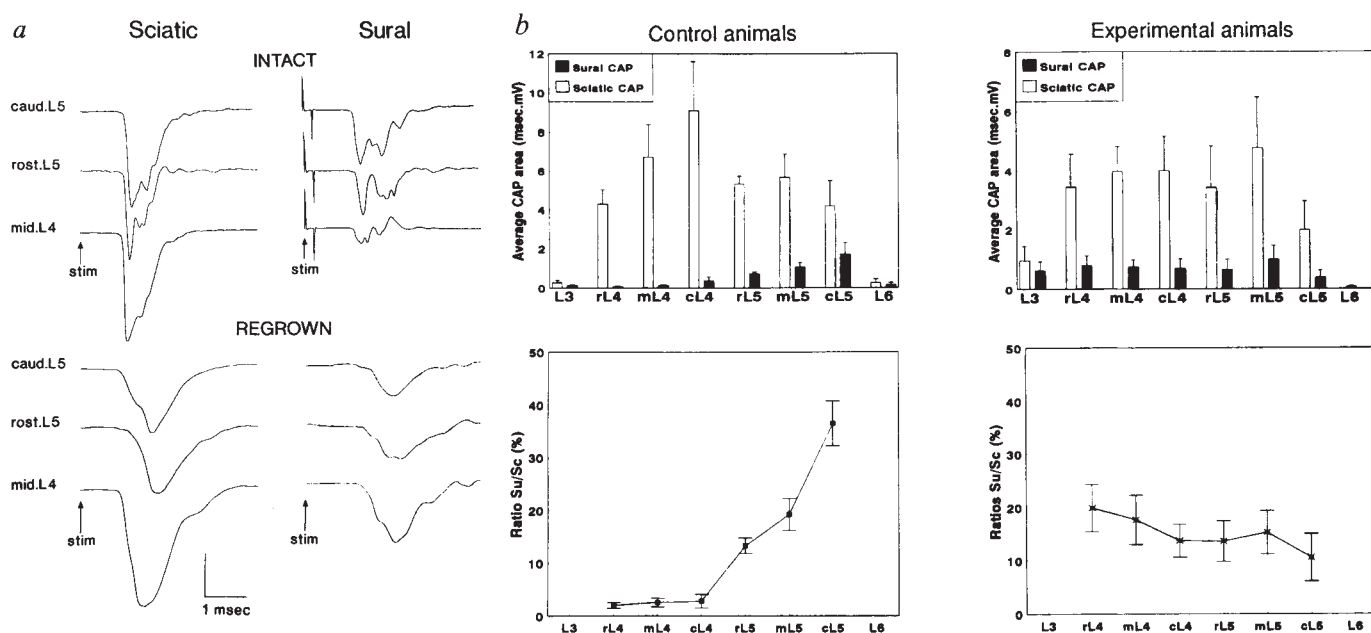


FIG. 4 *a*, Representative A-fibre compound action potentials (CAPs) recorded from dorsal rootlets following supramaximal A-fibre electrical stimulation of sciatic (shown on left) and sural (shown on right) nerves, in terminal experiments under urethane anaesthesia (1.25 g kg^{-1} ; i.p.). Upper panel shows responses following stimulation of intact nerves, lower panel shows responses after sciatic nerve entubation and regeneration (lesion made 12–14 weeks previously, as for Fig. 1). For these experiments, some dorsal roots were acutely subdivided into approximately equal-sized bundles, so that recording could be made from rostral L4, mid L4, caudal L4, rostral L5, mid L5, caudal L5 and the whole of L6. The potentials recorded from just three root bundles (caudal L5, rostral L5 and mid L4) are shown in these examples. Vertical calibration for intact and regenerated nerves is 1 mV and 0.5 mV,

respectively. *b*, Average CAP areas, recorded on different dorsal root bundles (as in *a*) and evoked from stimulation of sciatic nerve (open columns) and sural nerve (filled columns) in four normal animals (left) and four animals with regenerated sciatic nerves (right); error bars show s.e. Note that in the intact nerves, the size of the sciatic evoked CAPs was approximately equal across L4 and L5; the CAPs evoked from the sural nerve are larger in mid- and caudal L5 than in L4, as shown in the lower panel, where the ratio of sural:sciatic CAP area is plotted (slope of this line by linear regression is statistically different from 0; t value = 2.43, d.f. = 22, $P < 0.05$). After regeneration, this asymmetry is lost, and the projection to the sural nerve appears to be a random subset of sciatic afferents. The slope of the sural:sciatic CAP area is not statistically different from 0 ($P > 0.3$, t value = 0.94, d.f. = 22).

will be largely coincident during the animal's normal behaviour, and is therefore more likely to be associated with large depolarizations of postsynaptic target neurons. We propose that this allows NMDA receptors on the cell to mediate increases in synaptic efficacy of active inputs at the expense of non-coincidentally activated afferents²². Chronic blockade of NMDA receptors prevents dorsal horn neurons from 'selecting' only those afferents that have adjacent RFs. This model conforms to Hebb's rule²³, and has close parallels with the phenomenon of long-term potentiation in the hippocampus^{24,25} (also described in dorsal horn²⁶). It is possible that an anatomical rearrangement of

afferent terminals ('sprouting') could contribute to the changes observed.

An activity-dependent mechanism has been postulated to be responsible for the fine tuning of cortical somatosensory maps or their modification after peripheral nerve injury in adult^{27,28} and developing animals²⁹. Central neurons with small RFs are of course necessary for the generation of fine-grain somatosensory maps. Our results indicate that activity-dependent tuning of RF size at the first sensory relay site in the spinal cord may be a fundamental step in the maintenance and plasticity of somatosensory maps. □

Received 23 December 1993; accepted 27 April 1994.

1. Fawcett, J. W. & Keynes, R. J. A. *Rev. Neurosci.* **13**, 43–60 (1990).
2. Constantine-Paton, M., Cline, H. T. & Debeski, E. A. *Rev. Neurosci.* **13**, 129–154 (1990).
3. Dykes, R. W. & Terzis, J. K. *J. Neurophysiol.* **42**, 1461–1478 (1979).
4. Horch, K. *J. Neurophysiol.* **42**, 1437–1449 (1979).
5. Devor, M. & Wall, P. D. *Nature* **276**, 75–76 (1978).
6. Devor, M. & Wall, P. D. *J. Neurosci.* **1**, 679–684 (1981).
7. Lisney, S. J. W. *Brain Res.* **259**, 31–39 (1983).
8. McMahon, S. B., Lewin, G. R. & Wall, P. D. *Curr. Opin. Neurobiol.* **3**, 602–610 (1993).
9. Tulle, T. R., Berthele, A., Zieglansberger, W., Seeburg, P. H. & Wisden, W. *J. Neurosci.* **13**, 5009–5028 (1993).
10. Wong, E. H. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7104–7108 (1986).
11. Woolf, C. J. & Thompson, S. W. N. *Pain* **44**, 293–299 (1991).
12. Hao, J. X., Aldskogius, H., Seiger, A. & Weisenfeld-Hallin, Z. *Expl. Neurol.* **113**, 182–191 (1991).
13. Headley, P. M., Parsons, C. G. & West, D. C. *J. Physiol.* **385**, 169–180 (1984).
14. Dougherty, P. M., Palecek, J., Paleckova, V., Sorkin, L. S. & Willis, W. D. *J. Neurosci.* **12**, 3025–3041 (1992).
15. Davies, S. N. & Lodge, D. *Brain Res.* **424**, 402–406 (1987).
16. Dickenson, A. H. & Sullivan, A. F. *Neuropharmacology* **26**, 1235–1238 (1987).

17. Agrawal, S. G. & Evans, R. H. *Br. J. Pharmac.* **87**, 345–355 (1986).
18. Huettner, J. E. *Neuron* **5**, 255–266 (1990).
19. Li, Y., Erzurumlu, R. S., Chen, C., Jhaveri, S. & Tonegawa, S. *Cell* **76**, 427–437 (1994).
20. Thompson, S. W. N., King, A. E. & Woolf, C. J. *Eur. J. Neurosci.* **2**, 638–649 (1990).
21. Schaible, H.-G., Grubb, B. D., Neugebauer, V. & Oppmann, M. *Eur. J. Neurosci.* **3**, 981–991 (1991).
22. Lewin, G. R. & McMahon, S. B. *Eur. J. Neurosci.* **5**, 1083–1092 (1993).
23. Hebb, D. O. *The Organisation of Behaviour* (Wiley, New York, 1949).
24. Morris, R. G., Davis, S. & Butcher, S. P. *Phil. Trans. R. Soc. Lond.* **B329**, 187–204 (1990).
25. Watkins, J. C. *The NMDA Receptor Concept: Origins and Development* (eds Watkins, J. C. & Collingridge, G. L.) 1–19 (Oxford University Press, Oxford, 1989).
26. Randic, M., Jiang, M. C. & Cerne, R. J. *Neurosci.* **13**, 5228–5241 (1993).
27. Kaas, J. H. A. *Rev. Neurosci.* **14**, 137–167 (1991).
28. Kano, M., Lino, K. & Kano, M. *NeuroReport* **2**, 77–80 (1991).
29. Schlagger, B. L., Fox, K. & O'Leary, D. D. *Nature* **364**, 623–625 (1993).

ACKNOWLEDGEMENTS. This work was supported by a NATO twinning grant and the MRC. We thank L. Mendell for use of facilities and for critically reading the manuscript, C. Abel for technical assistance, and N. Taub for statistical advice.

Involvement of a calcineurin/inhibitor-1 phosphatase cascade in hippocampal long-term depression

Rosel M. Mulkey*, Shogo Endo†, Shirish Shenolikar† & Robert C. Malenka*‡

* Departments of Psychiatry & Physiology, University of California, San Francisco, California 94143-0984, USA

† Department of Pharmacology, Duke University Medical Center, Durham, North Carolina 27710, USA

‡ To whom correspondence should be addressed

LONG-TERM potentiation (LTP) is a synaptic mechanism thought to be involved in learning and memory¹. Long-term depression (LTD), an activity-dependent decrease in synaptic efficacy, may be an equally important mechanism which permits neural networks to store information more effectively^{2,3}. One form of LTD that has been observed in the hippocampus⁴ requires activation of post-synaptic NMDA (*N*-methyl-D-aspartate) receptors^{4,5}, a change in postsynaptic calcium concentration⁵, and activation of postsynaptic serine/threonine protein phosphatase 1 (PP1) or 2A (PP2A)⁶. The mechanism by which PP1 or PP2A is regulated by synaptic activity is unclear because these protein phosphatases are not directly influenced by calcium concentration. LTD induction may require activation of a more complex protein phosphatase cascade consisting of the Ca^{2+} /calmodulin-dependent protein phosphatase, calcineurin, its phosphoprotein substrate, inhibitor-1, and PP1^{7,8}. We tested this hypothesis using calcineurin inhibitors as well as different forms of inhibitor-1 loaded into postsynaptic cells. Our results suggest a signalling pathway in which calcineurin dephosphorylates and inactivates inhibitor-1. This in turn increases PP1 activity and contributes to the generation of LTD.

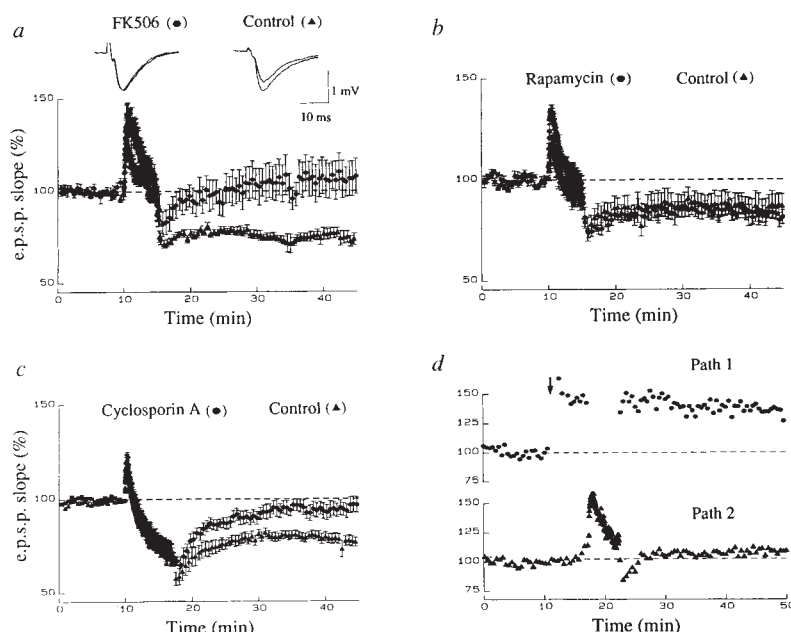
FIG. 1 Effects of extracellular application of calcineurin inhibitors on LTD and LTP. *a*, Summary of experiments in which the effects on LTD of application of FK506 (50–100 μM ; $n=8$) were compared with slices incubated in solvent (0.1–0.2% ethanol) alone ($n=9$). Data traces (average of 3 consecutive sweeps) show control responses and responses taken 30 min after LTD induction for a control (right traces) or an FK506-treated slice (left traces). *b*, Effects of rapamycin application (50 μM , $n=10$) compared with control slices ($n=6$) incubated in the solvent (1% ethanol). *c*, Effects of CsA application (250 μM , $n=8$) compared with control slices incubated in solvent (1% ethanol, 0.5% Tween 80) ($n=9$). *d*, Example of an experiment in which the effects of FK506 on LTP and LTD were examined in a single slice. Two independent paths were alternately stimulated. A 100 Hz/1 s tetanus to path 1 (arrow) generated LTP. At 17 min, 1 Hz stimulation for 5 min to path 2 did not generate LTD thus confirming the efficacy of the FK506 treatment.

METHODS. Hippocampal slices were prepared from Sprague-Dawley rats (age 12–20 days) using standard techniques. After a 1–2 h recovery period, slices were placed in oxygenated solutions containing either the experimental drug or the solvent alone. Slices remained in these solutions for 2–4 h at which time a single slice was transferred to a submersion type recording chamber⁶. Experiments on control slices were interleaved with those on experimental slices. Stimulating and recording procedures were as previously described⁶. LTD was induced by applying 1 Hz stimulation for 5–7 min^{4,5}. To compare the effects of FK506 on LTP and LTD in the same slice, two independent afferent fibre bundles were alternately stimulated (d , $n=$

The immunosuppressive drugs FK506 and cyclosporin A (CsA) are highly specific calcineurin inhibitors⁹. Incubation of slices in FK506 (50–100 μM) blocked LTD ($11 \pm 10\%$, $n=8$; mean per cent change from baseline $\pm$ s.e.m.) whereas control slices incubated in the solvent for FK506 (0.1–0.2% ethanol) had normal LTD ($-24 \pm 3\%$, $n=9$; $P<0.01$) (Fig. 1*a*). In contrast, application of rapamycin (50 μM), a compound with a structure similar to FK506 but lacking any calcineurin inhibitory activity⁹, had no significant effect on the ability to elicit LTD ($-17 \pm 3\%$, $n=10$) when compared to control slices incubated in the solvent alone (1% ethanol) ($-15 \pm 5\%$, $n=6$; $P>0.5$) (Fig. 1*b*). CsA (250 μM), a calcineurin inhibitor which requires an independent receptor protein⁹, also blocked the generation of LTD ($-4 \pm 4\%$, $n=8$) whereas control slices (1% ethanol and 0.5% Tween 80) had normal LTD ($-20 \pm 3\%$, $n=9$; $P<0.01$) (Fig. 1*c*).

Because LTP requires activation of protein kinases¹ and should not be blocked by protein phosphatase inhibitors, we compared the effects of FK506 on the ability to generate LTP and LTD in the same slice. FK506 had no discernible effect on the ability to generate LTP (Fig. 1*d*, top; $57 \pm 10\%$, $n=8$) and, consistent with the previous results (Fig. 1*a*), prolonged 1 Hz stimulation did not elicit LTD in these same slices (Fig. 1*d*, bottom; $13 \pm 4\%$, $n=8$).

A critical question which these experiments do not address is whether the calcineurin inhibitors were acting pre- or postsynaptically to block LTD. Therefore, we examined the effects of directly loading CA1 pyramidal cells with FK506 while simultaneously recording field excitatory postsynaptic potentials (e.p.s.ps) in each experiment to ensure that LTD was elicited in surrounding cells. There was a strong correspondence between the magnitude of LTD in control cells loaded with the solvent (0.1% ethanol) and the field e.p.s.ps (cells: $-32 \pm 6\%$; fields: $-26 \pm 2\%$, $n=13$) (Fig. 2*Aa, b*). In contrast, LTD was blocked in 5 of 7 cells loaded with FK506 (10–50 μM) and thus the mean magnitude of LTD was significantly different from the LTD elicited in surrounding cells (cells: $-7 \pm 5\%$; fields: $-31 \pm 3\%$, $n=7$, $P<0.01$) (Fig. 2*Aa, b*).



3) or LTD-inducing stimuli were given followed by LTP-inducing stimuli applied to the same afferent pathway ($n=5$). FK506 and rapamycin were gifts from Fujisawa Pharmaceutical and Wyeth-Ayerst Research, respectively, and were made up daily as stocks of 50 mM (FK506) or 5 mM (rapamycin) in ethanol. Cyclosporin A was a gift from Sandoz Pharmaceuticals and was made up daily as a stock of 17 mM in ethanol and Tween 80.

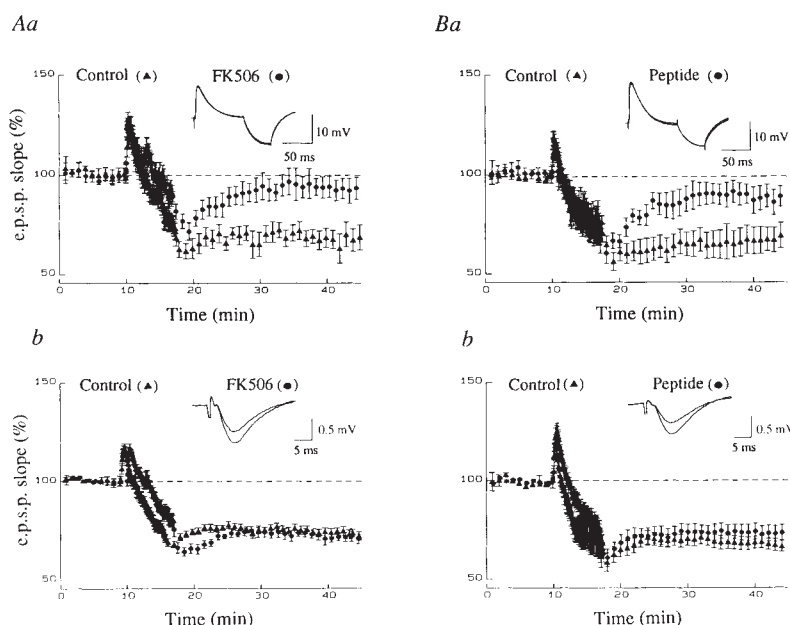
We also investigated the effects on LTD of loading cells with a third calcineurin inhibitor; a calcineurin autoinhibitory peptide (250–500 μ M), which blocks calcineurin activity through a mechanism distinct from that of FK506 or CsA¹⁰. LTD was blocked in 7 of 9 cells loaded with this peptide and this significantly reduced the magnitude of LTD in the cells ($-9 \pm 5\%$, $n=9$) (Fig. 2Ba) when compared with the LTD generated in the

field e.p.s.ps ($-36 \pm 3\%$, $n=9$; $P<0.001$) (Fig. 2Bb). There was again a strong correspondence between the magnitude of LTD in the control cells for the peptide experiments and the field e.p.s.ps (cells, $-33 \pm 7\%$, fields, $-27 \pm 4\%$; $n=12$) (Fig. 2Ba, b).

We have demonstrated that 3 different inhibitors of calcineurin block the generation of LTD. Inhibitor-1 (I1) when phosphorylated on threonine 35 inhibits PPI activity¹¹. Therefore

FIG. 2 Loading CA1 cells with calcineurin inhibitors blocks LTD. **Aa**, Summary of experiments in which the whole-cell pipettes used for recording e.p.s.ps were filled with a solution containing FK506 (10–50 μ M) ($n=7$) or the solvent (0.1% ethanol) alone ($n=13$). **Ab**, Graphs show the consequences of 1 Hz stimulation on the field e.p.s.ps which were recorded simultaneously during the experiments summarized in **a**. Data traces are averages of 3 consecutive sweeps obtained during the baseline period and 30 min after the induction of LTD and were taken from a typical experiment in which the recording pipette contained FK506. **Ba**, Summary of experiments in which the pipette solution contained the calcineurin autoinhibitory peptide¹⁰ (250–500 μ M) ($n=9$) or control solution alone ($n=12$). **Bb**, Graphs again show that in all experiments, LTD was generated in the simultaneously recorded field e.p.s.ps. Data traces as in **Aa** but taken from an experiment in which the pipette solution contained autoinhibitory peptide.

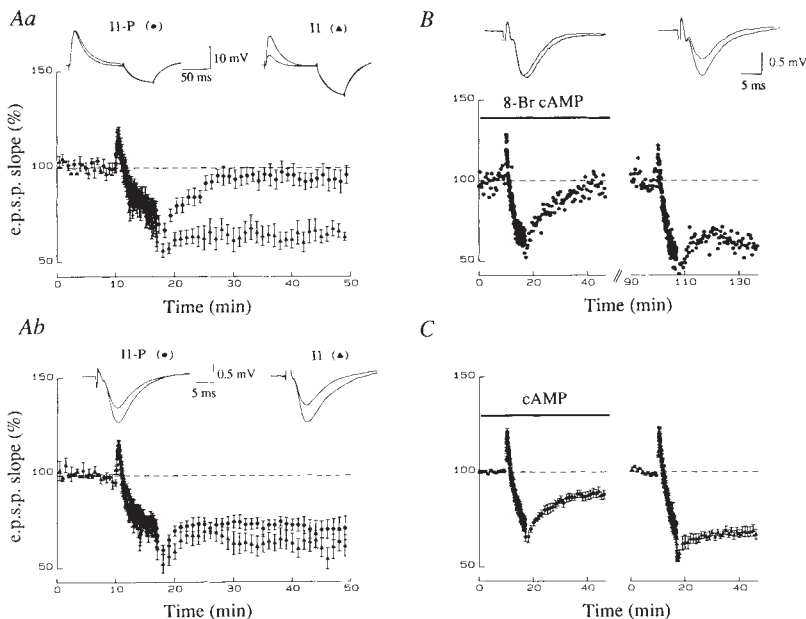
METHODS. Whole-cell recording techniques and pipette solutions are the same as previously described⁶. Experiments using control pipette solutions were always interleaved on a daily basis with those using calcineurin inhibitors. On each experimental day, the same batch of pipette solution was used in both the control and experimental cells. FK506 or the autoinhibitory peptide were added to the pipette solution immediately before the start of each experiment. To minimize the consequences of intracellular 'wash-out', 1 Hz stimulation was always applied within 15 min of break-in. All experiments in which, after 1 Hz stimulation, the field e.p.s.p. initial slope decreased by at least 15% were included in the final data analysis



(65 of 71 experiments). Combining the results from all control experiments in this study ($n=39$; the 8 additional cells not included in the figures were recorded with normal pipette solution) yields a significant correlation ($r=0.63$, $P<0.001$) between the magnitude of LTD obtained in the field e.p.s.p. ($-30 \pm 2\%$, $n=39$) and that in the individual cell ($-31 \pm 2\%$, $n=39$) using whole-cell recording techniques.

FIG. 3 Phosphorylation of inhibitor-1 blocks LTD. **Aa**, Summary of experiments in which the whole-cell pipettes were filled with a solution containing thiophosphorylated I1 (I1-P) ($n=10$) or unphosphorylated I1 (I1) ($n=7$). **Ab**, Summary of the changes in the field e.p.s.ps which were recorded simultaneously during the experiments in **a**. Data traces are averages of 3 consecutive sweeps obtained during the baseline period and 30 min after the induction of LTD. **B**, Example of an experiment in which application of 8-bromo-cAMP (1 mM) blocked the induction of LTD by 1 Hz/7 min stimulation. After a 50 min wash-out period, the same 1 Hz/7 min stimulation elicited LTD. Data traces were taken during the baseline period and 30 min after LTD induction. **C**, Summary of experiments ($n=13$) in which 1 Hz/7 min stimulation was given during application of either 8-bromo-cAMP ($n=6$) or Sp-cAMPS ($n=7$). The right graph shows the effects of 1 Hz/7 min stimulation in the same path after a 40–70 min wash-out period ($n=10$).

METHODS. I1 or I1-P was added to the pipette solution daily at a final concentration of 50 units μ l⁻¹ and experiments were as in Fig. 2. I1 was purified from rabbit skeletal muscle essentially according to ref. 12 with an addition of a final purification step on preparative SDS-PAGE (Biorad). This additional step was necessary because I1 purification by column chromatography yields a small contamination with PKA 'Walsh' inhibitor. This can be completely eliminated by SDS-PAGE. Thiophosphorylation of I1 blocks or strongly reduces the dephosphorylation of I1 by endogenous phosphatases¹⁷. For experiments shown in **B** and **C**, the adenosine receptor antagonist 8-cyclopentyl-1,3-dimethylxanthine (CPT; 10 μ M) was included in both



the control and experimental perfusing media. When necessary after wash-out of the cAMP analogue, stimulation strength was increased slightly to match e.p.s.p. slope to the original control value. The data in the right graphs (**B** and **C**) were normalized to the same original baseline as the data in the left graphs.

dephosphorylation of I1 by calcineurin leads to increased PP1 activity^{8,11}. To test if LTD induction is regulated by the phosphorylation state of I1, cells were loaded with I1 which had been thiophosphorylated on threonine 35¹². LTD was completely blocked in 7 of 10 cells which were loaded with thiophosphorylated I1 ($-6 \pm 5\%$, $n=10$), whereas LTD in the surrounding cells was normal ($-30 \pm 5\%$, $n=10$; $P<0.001$) (Fig. 3*Aa, b*). To test for effects of I1 unrelated to its inhibition of PP1, cells were loaded with unphosphorylated I1. This manipulation had no effect on LTD (Fig. 3*Aa, b*) (cells: $-36 \pm 5\%$, $n=7$; fields: $-38 \pm 5\%$).

These results suggest that phosphorylation of I1 by endogenous kinases may be an important regulatory step in LTD induction. Because the cyclic AMP-dependent protein kinase phosphorylates I1 on threonine 35¹¹, in preliminary work we examined the effects of forskolin on LTD and found that some LTD could still be elicited⁶. However, when membrane-permeable analogues of cAMP (Sp-cAMPS, 100 μ M, $n=7$ or 8-bromo-cAMP, 1 mM, $n=6$) were applied with the phosphodiesterase inhibitor IBMX (50 μ M), LTD was blocked in 10 of 13 slices. Thus the magnitude of LTD was reduced ($-12 \pm 2\%$, $n=13$) when compared to the LTD elicited in the same slices following wash-out of the cAMP analogue ($-32 \pm 3\%$, $n=10$; $P<0.01$) (Fig. 3*B, C*).

The present results demonstrate that a protein phosphatase cascade involving calcineurin, I1 and PP1 is probably required in the postsynaptic cell to induce LTD. This model (Fig. 4) (modified from ref. 8) is based on the regulation of PP1 activity by calcineurin and I1^{7,11} and proposes that a small Ca^{2+} influx through NMDA receptor-dependent channels is sufficient for activation of calcineurin by Ca^{2+} /calmodulin (CaM). Calcineurin then dephosphorylates I1, which in its phosphorylated state is a potent inhibitor of PP1. Thus activation of calcineurin results in an increase in PP1 activity by way of disinhibition. Both calcineurin and PP1 are found in high concentrations in brain¹³ with PP1 being highly concentrated in synaptic junctions¹⁴ and postsynaptic density fractions¹⁵. Furthermore, calcineurin has a high affinity for Ca^{2+} /CaM¹⁶ (dissociation constant, $K_d=0.1$ nM) and thus may be activated at much lower concentrations of Ca^{2+} /CaM than Ca^{2+} /CaM-dependent kinase II (CaMKII; $K_d=45$ nM)¹⁶, a protein kinase implicated in the generation of LTP¹.

Although a protein phosphatase cascade is involved in LTD, it is thought that a cascade of protein kinases may contribute to the induction of LTP¹. Thus the control of synaptic efficacy at these synapses may be under the regulation of a complicated

network of interacting and mutually regulatory signalling cascades, the functions of which are to control the phosphorylation state of critical substrate phosphoproteins. Which cascade dominates at any given instance may depend in part on the spatial and temporal dynamics of changes in postsynaptic calcium concentration which are profoundly influenced by synaptic activity. The critical substrate proteins contributing to the control of synaptic efficacy remain to be determined. □

Received 20 January; accepted 21 April 1994.

- Bliss, T. V. P. & Collingridge, G. A. *Nature* **361**, 31–39 (1993).
- Sejnowski, T. J. *J. math. Biol.* **4**, 303–321 (1977).
- Willshaw, D. & Dayan, P. *Neural Comp.* **2**, 85–93 (1990).
- Dudek, S. & Bear, M. F. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4363–4367 (1992).
- Mulkey, R. M. & Malenka, R. C. *Neuron* **9**, 967–975 (1992).
- Mulkey, R. M., Herron, C.H. & Malenka, R. C. *Science* **261**, 1051–1055 (1993).
- Cohen, P. *Proc. R. Soc., Lond.* **234**, 115–144 (1988).
- Lisman, J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9574–9578 (1989).
- Kunz, J. & Hall, M. N. *Trends biochem. Sci.* **18**, 334–338 (1993).
- Hashimoto, Y., Perrino, B. A. & Soderling, T. R. *J. Biol. Chem.* **265**, 1924–1927 (1990).
- Shenolikar, S. & Nairn, A. C. *Adv. Sec. Mess. Phosphoprot. Res.* **23**, 1–121 (1991).
- Shenolikar, S. & Ingebritsen, T. S. *Meth. Enzym.* **107**, 102–129 (1984).
- Stemmer, P. & Klee, C. B. *Curr. Opin. Neurobiol.* **1**, 53–64 (1991).
- Shields, S. M., Ingebritsen, T. S. & Kelly, P. T. *J. Neurosci.* **5**, 3414–3422 (1985).
- Dosemeci, A. & Reese, T. S. *J. Neurochem.* **61**, 550–555 (1993).
- Cohen, P. & Klee, C. B. (eds) *Calmodulin* (Elsevier, Amsterdam, 1988).
- Gratecos, D. & Fischer, E. H. *Biochem. biophys. Res. Commun.* **58**, 960–967 (1974).

ACKNOWLEDGEMENTS. This work was supported by grants from the Klingenstein Foundation, The McKnight Endowment Fund for Neuroscience, the National Alliance for Research on Schizophrenia and Depression, and the National Institute of Mental Health. R.C.M. is a member of the Center for Neurobiology and Psychiatry.

Altered microtubule organization in small-calibre axons of mice lacking tau protein

A. Harada*[†], K. Oguchi*, S. Okabe*, J. Kuno[†], S. Terada*, T. Ohshima*, R. Sato-Yoshitake*, Y. Takei*, T. Noda[‡] & N. Hirokawa*[§]

* Department of Anatomy and Cell Biology, School of Medicine, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

[†] Department of Neuropathology and Neuroscience, School of Pharmaceutical Sciences, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

[‡] Department of Cell Biology, Cancer Institute, Kamiikebukuro, Toshima-ku, Tokyo 170, Japan

THE *tau* gene encodes a protein (Tau) that is a major neuronal microtubule-associated protein localized mostly in axons^{1–4}. It has microtubule-binding and tubulin-polymerizing activity *in vitro*^{3,4} and is thought to make short crossbridges between axonal microtubules^{5,6}. Further, *tau*-transfected non-neuronal cells extend long axon-like processes in which microtubule bundles resembling those in axons are formed^{6–8}. In contrast, *tau* antisense oligonucleotides selectively suppress axonal elongation in cultured neurons^{9,10}. Thus *tau* is thought to be essential for neuronal cell morphogenesis, especially axonal elongation and maintenance. To test this hypothesis, we used gene targeting to produce mice lacking the *tau* gene. We show that the nervous system of *tau*-deficient mice appears to be normal immunohistologically. Furthermore, axonal elongation is not affected in cultured neurons. But in some small-calibre axons, microtubule stability is decreased and microtubule organization is significantly changed. We observed an increase in microtubule-associated protein 1A which may compensate for the functions of *tau* in large-calibre axons. Our results argue against the suggested role of *tau* in axonal elongation but confirm that it is crucial in the stabilization and organization of axonal microtubules in a certain type of axon.

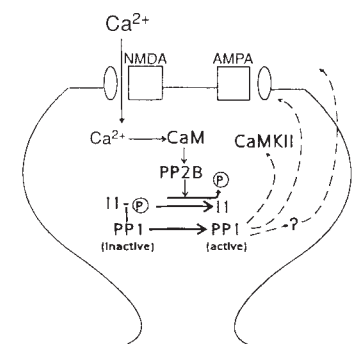


FIG. 4 Essential components of a protein phosphatase cascade which may mediate LTD. During prolonged 1 Hz stimulation, Ca^{2+} enters through the NMDA receptor-ionophore, binds to calmodulin (CaM) resulting in the activation of calcineurin (PP2B). Calcineurin then dephosphorylates I1 which therefore no longer inhibits PP1. The active PP1 may act on any number of substrates (dashed lines) including Ca^{2+} /calmodulin-dependent protein kinase II (CaMKII), non-NMDA receptor subunits, or some unknown process (?) which influences the production of putative retrograde messengers.

§ To whom correspondence should be addressed.

To disrupt the endogenous *tau* gene, a replacement-type targeting vector was constructed¹¹ (Fig. 1a). Genomic Southern blot analysis revealed a targeting frequency of about 1 in 5 G418^R clones and 1 in 2 G418^RFAU^R clones. Chimaeric male mice generated from injections with three different cell lines transmitted the embryonic stem (ES) cell genome through the germ line. When we interbred heterozygous animals and genotyped litters after weaning, the predicted 1:2:1 mendelian distribution of wild-type (*tau*^{+/+}), heterozygous (*tau*^{+/-}), and homozygous mutant (*tau*^{-/-}) animals was found (Fig. 1b). The *tau*^{-/-} animals were phenotypically indistinguishable at a gross level from their *tau*^{+/-} and *tau*^{+/+} littermates. Histological examination of the various tissues of *tau*^{-/-} mice by haematoxylin-eosin staining revealed no abnormalities in general throughout development.

Immunoblotting with various anti-Tau antibodies that recognize different epitopes of Tau did not detect Tau of either low (Fig. 1c) or high¹² relative molecular mass in crude extracts from *tau*^{-/-} brains (cerebrum and cerebellum) and dorsal root ganglion (DRG), whereas *tau* was readily detected in equivalent extracts from *tau*^{+/+} and *tau*^{+/-} littermates. Next we compared the level of each cytoskeletal protein in *tau*^{+/+}, *tau*^{+/-} and *tau*^{-/-} mice. The relative amounts of various microtubule-associated proteins (Tau, MAP1A, 1B, 2), neurofilament proteins, synapsin-1, and various tubulin isoforms (total α -, acetylated α -, tyrosinated α - and detyrosinated α -tubulin) were measured at postnatal day 7 (7d), 14 (14d), and adult (Ad). Of these

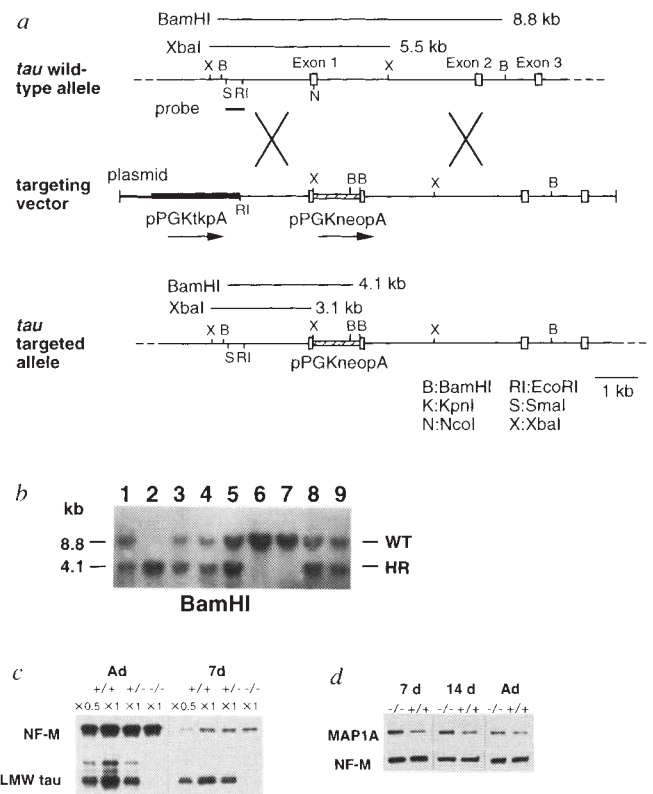
molecules, the amount of MAP1A was increased about 2-, 1.3- and 1.3-fold in 7d, 14d and adult *tau*^{-/-} mice, respectively (Fig. 1d), whereas the levels of other molecules in *tau*^{-/-} were similar to those in controls.

To assess possible minute structural defects in the nervous system of mutant mice, we stained frozen sections of cerebellar tissue from adult mice with antibodies against the same cytoskeletal proteins and with rhodamine-phalloidin. The staining with anti-Tau antibodies was prominently different in *tau*^{-/-} and control mice, whereas that with other antibodies was identical (data not shown). We obtained similar results using 7d and 14d cerebellum (data not shown).

To investigate further the morphological changes in *tau*^{-/-} brains, we examined cerebellar tissue by electron microscopy. Microtubule (MT) number and density in parallel fibre axons was reduced in *tau*^{-/-} mice (Fig. 2a, b). This reduction was statistically significant and was observed in two mouse lines derived from different ES cell clones (Table 1). This reduction was specific to axons, as the MT density in Purkinje cell dendrites was not significantly different between *tau*^{-/-} (MT density, 0.645 ± 0.052 (mean $\pm$ s.d.) μm^{-2} ; $n = 4$) and *tau*^{+/+} (MT density, 0.651 ± 0.010 μm^{-2} ; $n = 5$) sections. In addition, this observed reduction was not due to changes in axonal diameter because the average diameter of *tau*^{-/-} (116.7 ± 30.4 nm (mean $\pm$ s.d.); $n = 229$) and *tau*^{+/+} (118.4 ± 32.9 nm; $n = 532$) parallel fibre axons was not significantly different ($P > 0.5$; Student's *t*-test). We could not detect other morphological changes in parallel

FIG. 1 Targeted inactivation of the *tau* gene in ES cells and mice. a, Replacement-type targeting vector, *tau* wild-type allele, and targeted allele after homologous recombination. The targeting vector contains a PGK-neo cassette (pPGKneopA)²⁵ in the same transcriptional orientation as the *tau* gene and a PGK-tk cassette (pPGKtkpA) on the upstream side of the construct. As the *neo* cassette is inserted into the first coding exon of *tau*, only short fragments incapable of binding to MTs are expected to be produced from the mutated allele. The targeting vector was linearized with *SacI* before electroporation. b, Genotype of a litter from heterozygous $\times$ heterozygous interbreeding (*tau*^{+/-} $\times$ *tau*^{+/-}). Tail DNA was isolated and analysed by Southern blotting using the 5'-flanking probe ('probe' in a), after digesting with *Bam*HI. Lanes 6 and 7 are from wild-type (*tau*^{+/+}) mice. Lanes 1, 3, 4, 5, 8 and 9 are from heterozygous mutant (*tau*^{+/-}) mice. Lane 2 is from a homozygous mutant (*tau*^{-/-}) mouse. WT, wild-type band; HR, homologous recombinant band. c, d, Immunoblot analysis of MT-associated proteins. Crude extracts from brain (cerebrum plus cerebellum) were subjected to immunoblot analysis with antibodies against *tau* (c), MAP1A (d). c, Tau protein was reduced by half in *tau*^{+/-} and was absent in adult (Ad) and postnatal day 7 (7d) *tau*^{-/-} mice. d, MAP1A protein increased more in 7d, 14d and adult *tau*^{-/-} mice than in *tau*^{+/+} mice. Equivalent loadings of crude extract in each lane were confirmed by immunoblotting with an anti-NF-M antibody (see 'NF-M' in c, d).

METHODS. A 10-kb *Eco*RI fragment containing the *tau* gene was obtained from the 129/Sv mouse genomic library and subcloned into pGEM7Zf(+) (Promega). Exon numbering is according to ref. 23. The PGK-neo gene, derived from pKJ1²⁶, was blunt-end ligated in the *Nco*I site in the first coding exon in the same transcriptional orientation and PGK-tk was introduced into the upstream polylinker site. The J1 line of ES cells used for these experiments was cultured and electroporated essentially as described previously²⁶. ES cell colonies were expanded into cell lines and DNA was isolated for Southern blotting. Homologous recombination was confirmed using *Bam*HI with *Sma*I-*Eco*RI fragment ('probe' in a) and further confirmed using *Xba*I and *Eco*RV (not shown). ES cells carrying the disrupted *tau* gene were injected into C57BL/6 embryos at the blastocyst stage using standard techniques²⁷. As we obtained three lines of germ-line chimaeras, we bred and analysed two of them in detail and obtained similar results from either mouse line. Various parts of brain (cerebrum, cerebellum, spinal cord and DRG) were dissected and homogenized with 25 mM Tris, 2% SDS. Crude extracts were made by centrifuging the homogenates at 20,000g for 15 min at 4 °C. Immunoblotting was largely as described previously²⁸ except ¹²⁵I-labelled Protein A or ¹²⁵I-labelled anti-mouse IgG were used for the second antibodies. Binding was detected by autoradiography



using an Imaging Analyzer, Fuji BAS-2000 (Fuji-Film). For the first antibodies, the following monoclonal antibodies or polyclonal antisera were used: tau-1 (Boehringer Mannheim), anti-human *tau*²⁹ and anti-C4³⁰ for *tau*; HM2 (Sigma) for MAP2; 1B6 and 5E6¹⁹ for MAP1B; 1D1²⁰ for MAP1A; anti-NF-L antisera (TRI), NR4 (Sigma) for NF-L; anti-NF-M antisera (TRI), NN18 (Sigma) for NF-M; NE-14 (BioMakor) for NF-H; anti-synapsin 1 antisera²⁸ for synapsin 1; DM1A (Sigma) for α tubulin; 6-11B-1 (a gift from Sigma) for acetylated α -tubulin; SG (a gift from G. Gundersen) for detyrosinated α -tubulin; W2 (a gift from G. Gundersen) for tyrosinated α -tubulin.

TABLE 1 MT number and density in parallel fibre axons of homozygous mutant and control mice

Diameter†	50–99 nm			100–149 nm		
	No. of MTs	MT§ density	No. of axons	No. of MTs	MT density	No. of axons
T3-1‡						
+/+	2.00*	3.38	71	2.23	1.79	56
-/-	1.69	2.98	29	2.00	1.65	19
T3-2‡						
+/+	2.18***	3.34*	38	2.37***	1.86**	51
+/-	1.85*	2.98	26	2.41***	1.88**	27
-/-	1.54	2.72	39	1.67	1.36	33
U5‡						
+/+	2.04**	3.55**	23	2.20	1.67	20
+/-	1.61	2.91*	61	2.16	1.65	31
-/-	1.42	2.04	12	1.87	1.40	31

The number of MTs and MT densities in this table are averaged from axons whose numbers are indicated in the third columns (no. of axons). Student's *t*-test was used to determine the significance of the changes: * different from the values of -/- at $P < 0.05$; ** different from the values of -/- at $P < 0.01$; *** different from the values of -/- at $P < 0.001$.

† Axons were divided into two groups (50–99 nm and 100–149 nm) according to their shorter diameter.

‡ Mouse lines derived from two different ES cell clones (T3 and U5) were examined. Two sets of experiments (T3-1 or 2) used the T3-derived mouse line and one set (U5) used the U5-derived mouse line. In one set of experiments age-matched littermate mice were used.

§ MTs per 10^4 nm^2 .

fibre axons (such as axonal shape in cross-section). In the axons with larger diameter (optic nerve and sciatic nerve), MT density did not differ significantly (data not shown). This suggests that Tau affects the stability of MTs in small-calibre axons, which express *tau* proteins abundantly in controls.

To assess further the structural changes in axons of *tau*^{-/-} mice, axons were examined by quick-freeze, deep-etch (QF-DE) electron microscopy¹³. In *tau*^{+/+} cerebellar parallel fibre axons we frequently observed short filamentous crossbridges between MTs (Fig. 2c, white arrows). However in *tau*^{-/-} axons the frequency of crossbridges between MTs was greatly reduced. We often observed two MTs running parallel to each other for some distance with no crossbridges between them (Fig. 2d, long white arrows). This structure was never observed in *tau*^{+/+} axons. When we measured the frequency of crossbridges on MTs (number of crossbridges per 100 nm of MT), we found it greatly reduced in *tau*^{-/-} axons (4.00 ± 2.93 in *tau*^{-/-} (mean $\pm$ s.d.); $n = 10$ and 16.35 ± 4.06 in *tau*^{+/+}; $n = 10$) and this reduction was statistically significant ($P < 0.001$; Student's *t*-test). However, there were filamentous crossbridges between MT and the axonal plasma membrane in *tau*^{-/-} axons as well as in *tau*^{+/+} axons (Fig. 2c–e, arrows). In rare cases a neurofilament (Fig. 2f, long white arrow) ran parallel to MTs, when we observed short crossbridges between them (Fig. 2f, arrowheads). We conclude that Tau is the major component of the short crossbridges between MTs and that it controls MT organization in the axons.

To examine the effect of Tau in axonal elongation, we cultured hippocampal neurons from *tau*^{-/-} and control (*tau*^{+/+}) littermate embryos (Fig. 3a, b). The *tau*^{-/-} neurons appeared to be indistinguishable from control neurons in culture at all time points (at days 0.5, 1, 1.5, 2 and 3 after plating). We followed the cells photographically or by video-recordings and measured their fraction at stage 3 of development (stage of axonal elongation) (Fig. 3c)^{14,15}. Mutant neurons at stage 3 increased in number in the same way as control neurons in culture. When we cultured neurons in nocodazole (MT-depolymerizing drug), the results were identical (data not shown). We obtained similar results with DRG neurons (data not shown). This suggested that Tau is not essential for axonal elongation and neuronal polarization.

To investigate MT turnover in axons of living neurons, we measured fluorescence decay after photoactivation of caged

fluorescein-labelled tubulin^{16–18}. We microinjected caged fluorescein-labelled tubulin into cultured DRG neurons from *tau*^{-/-} and *tau*^{+/+} mice, photoactivated a part of the axons, and measured the fluorescence intensities. We did not observe any significant difference in the fluorescent decay between *tau*^{-/-} and

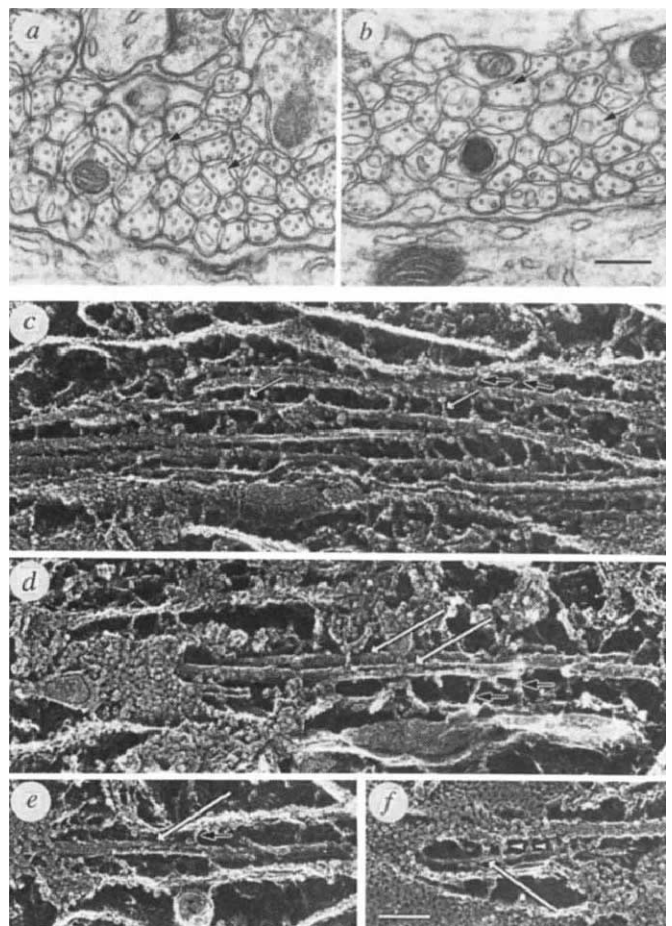


FIG. 2 a, b, Decreased MT density in *tau*^{-/-} parallel fibre axons. Electron micrographs compare representative areas of molecular layer in cerebellum of *tau*^{+/+} (a) and *tau*^{-/-} (b) animals. Arrows indicate MTs in parallel fibre axons. The MT density was reduced in *tau*^{-/-} axons compared to *tau*^{+/+} axons. Scale bar, 300 nm. c–f, Quick-frozen, deep-etched parallel fibre axons of a *tau*^{+/+} and a *tau*^{-/-} mouse. c, A parallel fibre of a *tau*^{+/+} mouse. A few MTs run parallel to each other in the axon, and numerous short crossbridges between them are apparent (white arrows). Crossbridges between MTs and axonal plasma membrane are also apparent (black arrows). d–f, Parallel fibres of a *tau*^{-/-} mouse. d, Two MTs (long white arrows) run parallel to each other for some distance without short crossbridges. Note that the crossbridges between MTs and axonal plasma membrane are still apparent (black arrows). e, Axons that contain only one MT are frequently observed (long white arrow). The crossbridges between a microtubule and the plasma membrane are also apparent (black arrow). f, MT and neurofilament (long white arrow) running parallel in the axon. Crossbridges between the microtubule and neurofilament are evident (arrowheads). Scale bar, 100 nm.

METHODS. Anaesthetized mice were perfused with 2% paraformaldehyde and 2.5% glutaraldehyde in 0.1 M cacodylate buffer. Tissues (cerebellum, optic nerve and sciatic nerve) were dissected out and further fixed overnight. Sections were processed and viewed as described²⁸. For analysis of axonal MT number and density, areas adjacent to Purkinje dendrites with well preserved MTs were selected from the cerebellar molecular layer, optic nerve and sciatic nerve of *tau*^{-/-} and age-matched control (*tau*^{+/+} and *tau*^{+/+}) littermates. MT cross-sections were counted using a blind-test protocol. For QF-DE electron microscopy, cerebellar tissues were dissected out, and thin slices of tissue were saponin-treated and processed as described previously^{5,13,28}.

$\tau^{+/+}$ DRG at both 20 and 40 min after activation (Fig. 3d). With a few samples, we measured the intensity at 60 min and again found no difference (data not shown). Our results indicated that τ is not crucial for MT stabilization in DRG neurites.

In conclusion, Tau is dispensable for axonal elongation but is important for MT organization in cerebellar parallel fibres. As for MT stability, we obtained two seemingly different results. On the one hand we did not observe significant differences in

MT stability. These results were derived from the observation of (1) MT density in optic and sciatic nerves, (2) the photoactivation study with DRG neurons, and (3) axonal morphology in cultured hippocampal neurons in the presence of nocodazole. On the other hand, we observed a decrease in MT number and density in parallel fibre axons whereas the total amount of tubulin was unchanged, which implies MTs tend to depolymerize in these axons. These two apparently contradictory results seem to reflect the relative amount of Tau expressed in small- and large-calibre axons. In large-calibre axons (optic and sciatic nerves and DRG and hippocampal axons), large amounts of other MT-associated proteins (such as MAP1A and 1B) are present, whereas the amount of Tau is relatively small^{19,21}. Thus the loss of Tau may hardly affect MT stability in these axons. The increased MAP1A may also compensate for the loss of Tau in these axons. However, in small-calibre axons where Tau occupies a large proportion of total MT-associated proteins, the loss of Tau would greatly affect the stability of MTs. As for axonal elongation, our result seems to contradict previous results from antisense treatment^{9,10}. A similar contradiction between these two methods has previously been reported²². Although the reason is still unclear, in our case a possible explanation may be that antisense inhibits some pathway(s) important for axonal elongation rather than solely inhibiting Tau synthesis. The sequence of the τ gene^{7,23} is well conserved among many species²⁴, which is indicative of its importance. Its functions might be revealed by examining axonal regeneration or various neural functions. Because we found that MT stability was decreased in the parallel fibre, which could lead to instability of the axonal structure of $\tau^{-/-}$ cerebellum, we first examined the neurological signs derived from cerebellar abnormalities. But we could detect no such abnormalities in our examination (data not shown). We are doing other neurological tests (sensory and motor system, memory) to find any other neurological abnormalities. Finally, some functional redundancy might exist between other MT-associated proteins. This possibility will soon be tested by intercrosses between Tau-deficient mice and mice deficient in other MT-associated proteins. □

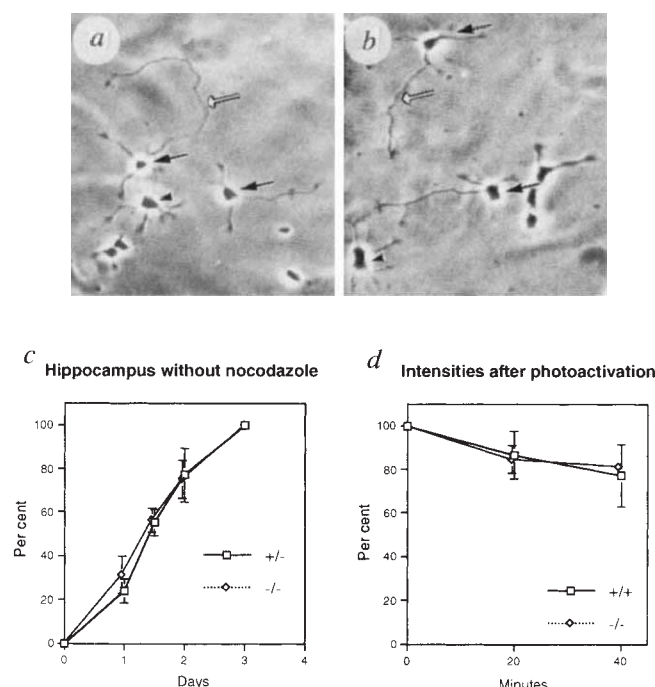


FIG. 3 Axonal elongation of hippocampal neurons. *a, b*, Phase micrographs of hippocampal cells in culture at 1.0 day after plating. *a*, Control ($\tau^{+/+}$) cells. *b*, $\tau^{-/-}$ cells. Axons are indicated by white arrows. Some cells are at stage 3 of development (stage of axonal elongation) (black arrows). Some stage 2 (stage of minor processes) cells are also present (arrowheads). Note that there is no significant difference in cell polarization between *a* and *b*. *c*, Hippocampal neurons from $\tau^{+/+}$ ($n=3$; mean $\pm$ s.d.; $\square$) and $\tau^{-/-}$ ($n=3$; $\diamond$) mice were photographed and the percentage of cells in stage 3 was measured. *d*, Time course of fluorescence intensity after photoactivation. Percentage of fluorescence intensity after activation in bars of fluorescence in $\tau^{+/+}$ cells ($n=11$ in 0, 20 min; $n=3$ in 40 min; mean $\pm$ s.d.; $\square$) and that in bars of fluorescence in $\tau^{-/-}$ cells ($n=10$ in 0, 20 min; $n=2$ in 40 min; $\diamond$). Fluorescence decay was similar regardless of the presence of τ .

METHODS. Methods for preparing the hippocampal cell cultures followed those previously described^{14,15} with the exception that the brains of 16.5-day-old mouse fetuses were used. Fetuses were collected from pregnant $\tau^{+/+}$ female mice mated with $\tau^{-/-}$ male mice. The fetuses were genotyped individually by genomic Southern blot after plating. In one set of experiments, fetuses belonging to the same litter and the same conditioned medium was used, and feeder glial cells were dissociated from littermate embryos. We followed the neurons photographically or by video-recordings using a similar protocol to ref. 14. Both recording methods yielded similar results. The photoactivation procedure was as described previously¹⁶⁻¹⁸. To minimize bias resulting from the experimental conditions used, $\tau^{-/-}$ and age-matched $\tau^{+/+}$ mice were chosen from littermates, DRG neurons were taken and dissociated using the same solution at the same time, and they were cultured on laminin-coated substrate prepared from the same batch. Photoactivation runs were done alternately from $\tau^{-/-}$ and $\tau^{+/+}$ DRG neurons during the period from 10 h to 24 h after microinjection. The photoactivated region was set to the proximal neurite $<100 \mu\text{m}$ from the cell body to eliminate the possibility of any regional difference of MT turnover along the axon. Image quantification was described in detail elsewhere¹⁸.

Received 13 December 1993; accepted 13 April 1994.

- Olmsted, J. B. *Rev. Cell Biol.* **2**, 421-457 (1986).
- Matus, A. A. *Rev. Neurosci.* **11**, 20-44 (1988).
- Cleveland, D. W., Hwo, S. Y. & Kirschner, M. W. *J. molec. Biol.* **116**, 207-225 (1977).
- Binder, L. I., Frankfurter, A. & Rebhun, L. I. *J. Cell Biol.* **101**, 1371-1378 (1985).
- Hirokawa, N., Shiomura, Y. & Okabe, S. *J. Cell Biol.* **107**, 1449-1461 (1988).
- Chen, J., Kanai, Y., Cowan, N. J. & Hirokawa, N. *Nature* **360**, 674-677 (1992).
- Chen, J., Kanai, Y., Liem, R. K. H., Mellado, W., Nunez, J. & Shelanski, M. L. *J. Cell Sci.* **100**, 55-60 (1991).
- Knops, J. et al. *J. Cell Biol.* **114**, 725-733 (1991).
- Caceres, A. & Kosik, K. S. *Nature* **343**, 461-463 (1990).
- Caceres, A., Potrebic, S. & Kosik, K. S. *J. Neurosci.* **11**, 1515-1523 (1991).
- Thomas, K. R. & Capocchi, M. R. *Cell* **51**, 503-512 (1987).
- Georgieff, I. S., Liem, R. K. H., Mellado, W., Nunez, J. & Shelanski, M. L. *J. Cell Sci.* **100**, 55-60 (1991).
- Hirokawa, N. *J. Cell Biol.* **94**, 129-142 (1982).
- Dotti, C. G., Sullivan, C. A. & Banker, G. A. *J. Neurosci.* **8**, 1454-1468 (1988).
- Goslin, K. & Banker, G. in *Culturing Nerve Cells* (eds Banker, G. & Goslin, K.) 251-282 (MIT Press, Cambridge, MA, 1991).
- Mitchison, T. J. *J. Cell Biol.* **109**, 637-652 (1989).
- Okabe, S. & Hirokawa, N. *J. Cell Biol.* **117**, 105-120 (1992).
- Uneyama, T., Okabe, S., Kanai, Y. & Hirokawa, N. *J. Cell Biol.* **120**, 451-465 (1993).
- Sato-Yoshitake, R., Shiomura, Y., Miyasaka, H. & Hirokawa, N. *Neuron* **3**, 229-238 (1989).
- Shiomura, Y. & Hirokawa, N. *J. Neurosci.* **7**, 1461-1469 (1987).
- Bloom, G. S., Schoenfeld, T. A. & Vallee, R. B. *J. Cell Biol.* **98**, 320-330 (1985).
- Lee, K. et al. *Cell* **69**, 737-749 (1992).
- Himmeler, A. *Molec. cell. Biol.* **9**, 1389-1396 (1989).
- Lee, G. *Cell Motil. Cytoskel.* **15**, 199-203 (1990).
- McBurney, M. W. et al. *Nucleic Acids Res.* **19**, 5755-5761 (1991).
- Li, E., Bestor, T. H. & Jaenisch, R. *Cell* **69**, 915-926 (1992).
- Bradley, A. in *Teratomas and Embryonic Stem Cells: A Practical Approach* (ed Robertson, E. J.) 113-152 (IRL, Oxford, 1987).
- Harada, A., Sobue, K. & Hirokawa, N. *Cell Struct. Funct.* **15**, 329-342 (1990).
- Ihara, Y. et al. *Gerontology* **36**, 15-24 (1990).
- Kondo, J. et al. *Neuron* **1**, 827-834 (1988).

ACKNOWLEDGEMENTS. We thank S. Ikegami, T. Ishikawa, H. Oda, K. Yamamura, H. Niwa, M. Fujiwara, G. Banker and H. Asmussen for help and discussion. This work was supported by a grant for special project research by the Ministry of Education, Science and Culture of Japan and a grant from RIKEN to N.H. and by a grant for scientific research to A.H.

A universal nucleoside for use at ambiguous sites in DNA primers

R. Nichols^{*†}, P. C. Andrews^{*}, P. Zhang^{‡§}
& D. E. Bergstrom^{‡§||}

^{*} Departments of Biological Chemistry and [†] Biology, University of Michigan, Ann Arbor, Michigan 48109, USA
[‡] Walther Cancer Institute, Indianapolis, Indiana 46208, USA
[§] Department of Medicinal Chemistry and Pharmacognosy, Hansen Life Sciences Research Building, Purdue University, West Lafayette, Indiana 47907, USA
^{||} To whom correspondence should be addressed at Purdue University

A NON-DISCRIMINATORY base analogue, or universal base, would be an invaluable component of oligonucleotide probes and primers for solving the design problems that arise as a result of the degeneracy of the genetic code, or when only fragmentary peptide sequence data are available. We have designed an alternative to previous universal nucleoside candidates¹⁻⁹, a new analogue, 1-(2'-deoxy- β -D-ribofuranosyl)-3-nitropyrrole (designated M; Fig. 1), which maximizes stacking while minimizing hydrogen-bonding interactions without sterically disrupting a DNA duplex. Oligonucleotides containing M at several sites were used as primers for sequencing and the polymerase chain reaction. The sequencing primer d(5'-CGT AAM CAM AAM ACM AT-3') is as effective as the exact match d(5'-CGT AAT CAG AAA ACA AT-3'). It is also possible to sequence using a primer containing M at several contiguous positions, for example d(5'-CGT AAT MMM MMM MMM AT-3'). Melting curves show that duplexes formed on hybridization of the sequences d(5'-CCT TTT TMT TTT TGG-3') and d(5'-CCA AAA AXA AAA AGG-3'), where X is A, C, G or T, melted at a lower temperature than the corresponding duplexes containing only d(A·T) and d(C·G) base pairs, but showing little variation among different X bases (T_m range 3 °C).

In our design and synthesis of nucleoside analogues, we aimed to maximize stacking interactions using aprotic polar substituents linked to heteroaromatic rings. Enhancing intra- and inter-strand stacking interactions should lessen the role of hydrogen bonding in base-pairing specificity. Functional groups that would be expected to polarize a π -aromatic system and thereby enhance stacking, but which are not strong hydrogen-bond acceptors, include cyano and nitro groups. For maximum polarity and stacking we chose 3-substituted pyrrole deoxyribonucleosides as the best candidates for a first generation of weakly hydrogen-bonding universal nucleosides. Of the candidate molecules, we favoured 3-nitropyrrole 2'-deoxyribonucleoside (M; Fig. 1) because of its structural and electronic resemblance to *p*-nitroaniline, whose derivatives are among the smallest known intercalators of double-stranded DNA¹⁰. 3-Nitropyrrole 2'-de-

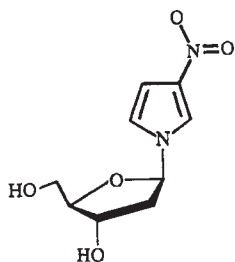


FIG. 1 Structure of 1-(2'-deoxy- β -D-ribofuranosyl)-3-nitropyrrole. Details of the synthesis and crystal structure coordinates will be published elsewhere (D.E.B., P.Z., P.H.T., P.C.A. and R.N., manuscript in preparation).

oxy-ribonucleoside was synthesized and its structure confirmed by X-ray crystallography. The dimethoxytrityl-protected phosphoramidite of nucleoside M was synthesized for incorporation into oligonucleotides used as primers for sequencing and polymerase chain reaction (PCR).

Sequencing studies show that a substantial number of nucleotides can be replaced by M without loss of primer specificity (Fig. 2a and b). Sequencing primer 4 with substitutions of M at the third position in each of four codons gave a sequencing ladder comparable to primer 1, the exact match, whereas a 256-fold degenerate oligonucleotide mixture (primer 2) gave an unreadable sequencing ladder (Fig. 2a). Comparison of primers 3 and 4 is instructive as the former contains four deoxyinosines at the same sites occupied by M in primer 4, and the sequencing ladder with primer 3 is only partially readable (Fig. 2a). In comparison, primers with two or more mismatches at the same

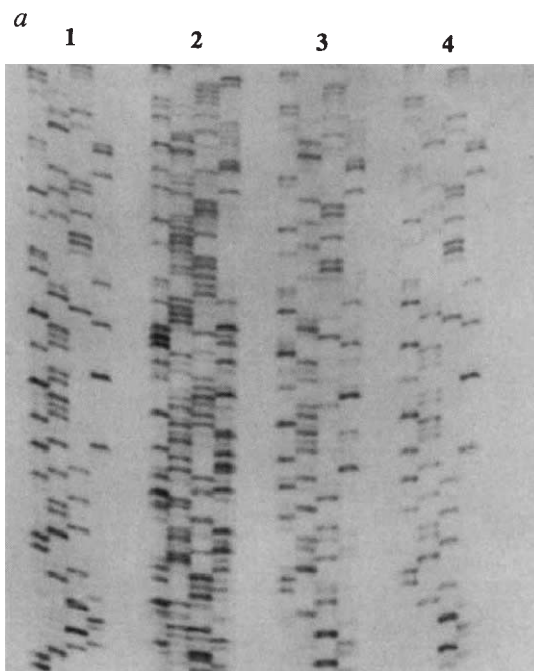


FIG. 2 a, Dideoxy sequencing using primers 1-4. Numbers correspond to the oligonucleotide sequencing primers d(5'-CGT AAT CAG AAA ACA AT-3') (primer 1); d(5'-CGT AAN CAN AAN ACN AT-3'), where N is A, C, G or T (primer 2); d(5'-CGT AAI CAI AAI ACI AT-3') (primer 3); and d(5'-CGT AAM CAM AAM ACM AT-3') (primer 4). The oligonucleotides were purified by HPLC on a 25 cm $\times$ 4 mm polyhydroxyethyl aspartamide column (polyLC) or by denaturing-gel electrophoresis and size-exclusion chromatography. For sequencing, Sequenase version 2.0 (USB) was used according to the manufacturer's recommendations. Single-stranded DNA (1 μ g) from *Drosophila melanogaster* sulfakinin¹⁴ in Bluescript SK⁺ (Stratagene) was sequenced using 0.1 μ g primer and ³⁵S-dATP (Amersham). Aliquots of sequencing reactions were electrophoresed on 6% acrylamide-8M urea gels, and the gels dried and exposed to Kodak XAR X-ray film. b, Dideoxy sequencing using primers 1 and 5-7. Numbers correspond to the oligonucleotide sequencing primers d(5'-CGT AAT CAG AAA MMM AT-3') (primer 5); d(5'-CGT AAT CAG MMM MMM AT-3') (primer 6); d(5'-CGT AAT MMM MMM MMM AT-3') (primer 7). Oligonucleotides were purified and sequenced as in a. The sequence ladder from primer 7 was consistently fainter than for primers 1, 5 and 6, resulting in overexposure of the first three lanes. c, PCR amplification using primers 1, 4, 8 and 9. Lane numbers correspond to the oligonucleotide primers. The sequence of primer 8 is d(5'-CGT AAT CAG AAA ACA AM-3'), and primer 9 is d(5'-CGT AAT CAG AAA ACM AT-3'). The lower band in each lane arises from primer dimerization. The expected product is ~315 bases long. The identity of the product was verified by dideoxy sequence analysis (data not shown). First-strand synthesis was done using total RNA extracted from *Drosophila melanogaster* heads and avian myeloblastosis virus reverse transcriptase (Promega) and

sites gave uninterpretable results. A unique property of M is its ability to replace long strings of contiguous nucleosides and still yield functional sequencing primers. Sequences with three (primer 5), six (primer 6) and nine (primer 7) M substitutions (Fig. 2b) all gave readable sequencing ladders. PCR with primer 1 and three different M-containing primers (primers 4, 8, 9) all resulted in amplification of the correct product (Fig. 2c).

The ability of 3-nitropyrrole-containing oligonucleotides to function as primers strongly suggests that a duplex structure must form with complementary strands. Optical thermal profiles obtained for the oligonucleotide pairs d(5'-C₂T₅XT₅G₂-3') and d(5'-C₂A₅YA₅G₂-3') (where X and Y can be A, C, G, T or M) fit the normal sigmoidal pattern observed for the DNA double-to-single strand transition. The melting temperatures (T_m) were determined at 260 nm in 1.0 M NaCl, 0.1 mM EDTA, 10 mM sodium phosphate, pH 7.0, at an oligonucleotide concentration of 50 μ M. The T_m values of the oligonucleotides containing X·M

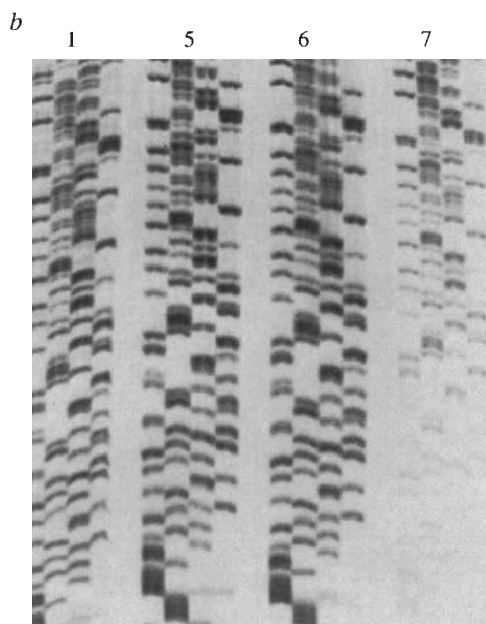
base pairs (where X was A, C, G or T, and Y was M) all fell within a 3 °C range. Deoxyinosine (dI), an extensively studied nucleoside with some characteristics of a universal nucleoside¹¹, is more discriminating. Duplexes containing dI opposite any base vary in T_m by as much as 15 °C, corresponding to differences in base-pair stability of 2–3 kcal mol⁻¹ (refs 12, 13).

This letter describes the first of a series of modified nucleosides which, in addition to their application as tools for use in molecular biology, should refine our understanding of the forces involved in DNA duplex formation and of the factors influencing specificity. □

Received 3 December 1993; accepted 25 April 1994.

1. Millican, T. A. et al. *Nucleic Acids Res.* **12**, 7435–7453 (1984).
2. Inoue, H., Imura, A. & Ohtsuka, E. *Nucleic Acids Res.* **13**, 7119–7128 (1985).
3. Fukuda, T., Hamana, T., Kikuchi, K. & Marumoto, R. Z. *Naturforsch. B.* **41**, 1571–1579 (1986).
4. Seela, F. & Kaiser, K. *Nucleic Acids Res.* **14**, 1825–1844 (1986).
5. Eritja, R. et al. *Nucleic Acids Res.* **14**, 8135–8153 (1986).
6. Habener, J. F. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1735–1739 (1988).
7. Lin, P. K. T. & Brown, D. M. *Nucleic Acids Res.* **17**, 10373–10383 (1989).
8. François, P., Perilleux, D., Kempener, Y. & Sonveaux, E. *Tetrahedron Lett.* **31**, 6347–6350 (1990).
9. Brown, D. M. & Lin, P. K. T. *Carbohydrate Res.* **216**, 129–139 (1991).
10. Gabbay, E. J. in *Bioorganic Chemistry* (ed. van Tameelen, E. E.) Vol. 3, 33–70 (Academic, New York, 1977).
11. Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y. & Matsubara, K. *J. biol. Chem.* **260**, 2605–2608 (1985).
12. Kawase, Y., Iwai, S., Inoue, H., Miura, K. & Ohtsuka, E. *Nucleic Acids Res.* **14**, 7727–7736 (1986).
13. Martin, F. H., Castro, M. M., Aboul-ela, F. & Tinoco, I. Jr *Nucleic Acids Res.* **13**, 8927–8938 (1985).
14. Nichols, R., Schneuwly, S. A. & Dixon, J. E. *J. biol. Chem.* **263**, 12167–12169 (1988).

ACKNOWLEDGEMENTS. We thank M. F. Tibbetts for sequencing, H. Liu for PCR, and S. E. Olsen and T. A. Polley for T_m studies. This work was supported by the NIH.



oligo(dT) at 42° C for ~1 h. The first-strand synthesis reaction was then divided into four aliquots and used for amplification with four different primers: 1, 4, 8 and 9. Each PCR was run for 30 cycles, with annealing at 48 °C, extension at 72 °C, and denaturing at 95 °C, using sequence-specific primer (primer 1, 4, 8 or 9), oligo(dT) and Taq polymerase (Perkin-Elmer) according to the manufacturer's suggestions. Each reaction was analysed by gel electrophoresis and visualized by ethidium bromide staining.

The role of Rab3A in neurotransmitter release

Martin Geppert^{*†}, Vadim Y. Bolshakov[‡],
Steven A. Siegelbaum[‡], Kohji Takei[§],
Pietro De Camilli[§], Robert E. Hammer[†]||
& Thomas C. Südhof^{*†¶}

^{*} Departments of Molecular Genetics and

^{||} Biochemistry and [†] Howard Hughes Medical Institute, The University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA

[‡] Department of Pharmacology, Center for Neurobiology and Behavior, Howard Hughes Medical Institute, Columbia University, New York, New York 10032, USA

[§] Department of Cell Biology and Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06510, USA

THE small GTP-binding protein Rab3A is a Rab family member^{1,3} that is abundant in brain synaptic vesicles^{4,5}. Here we show that mice in which the *rab3A* gene has been mutated by homologous recombination do not express Rab3A but are viable and fertile. Electrophysiological recordings in hippocampal CA1 pyramidal cells indicate that most of their synaptic parameters are also normal, although synaptic depression after short trains of repetitive stimuli (15–30 stimuli at 14 Hz) is significantly increased. Levels of the Rab3A-binding protein rabphilin are decreased by 70%, but expression of more than 20 other synaptic proteins is unchanged. No compensatory changes were detected in other GTP-binding proteins or in proteins that interact with Rab3. Rab3A thus appears not to be essential for synaptic vesicle exocytosis but to play a role in the recruitment of synaptic vesicles for exocytosis during repetitive stimulation.

^{*} To whom correspondence should be addressed.

Using clones of the murine *rab3A* gene⁶, we constructed a targeting vector to delete the promoter and the first two exons of the *rab3A* gene by homologous recombination²⁻⁹ (Fig. 1A). The linearized targeting vector was electroporated into embryonic stem (ES) cells and transfectants resistant to G418 and FIAU (1-[2-deoxy-2-fluoro- β -D-arabinofuranosyl]-5-iodouracil) were selected. Two hundred and two double-resistant clones were screened by polymerase chain reaction (PCR) and two clones with mutant *rab3A* genes were isolated. On injection into blastocysts, one of the clones produced chimaeric mice that transmitted the mutant gene through the germ line. Heterozygous mouse lines were established from independent chimaeras and bred to homozygosity. Immunoblotting showed that *rab3A* expression is completely abolished in homozygous mutant mice (Fig. 1B). Nevertheless, the mutant mice were viable, capable of mating and caring for offspring, and were comparable in weight and health to their littermates and to mice from control lines. No major changes in the histological features, synaptic connectivity or fine structure of their brains were visible by light or electron microscopy, suggesting that their neural development and synaptic structure are normal (Fig. 1C, and data not shown). Thus, Rab3A is not required for the fundamental operation of the nervous system.

To investigate the effects of Rab3A deficiency further, synaptic transmission was studied in wild-type and mutant mice. The size of excitatory postsynaptic currents (e.p.s.cs) in pyramidal cells from the CA1 region of the hippocampus upon stimulation of the Schaffer collateral-commissural fibres was unchanged, as were several forms of synaptic plasticity, including paired-pulse facilitation, post-tetanic potentiation and long-term potentiation^{10,11} (Fig. 2).

A striking characteristic of the synaptic vesicle pathway is the efficiency of membrane trafficking in the nerve terminal. Might Rab3A ensure efficient synaptic vesicle trafficking, a function of sufficient importance to account for its evolutionary conservation? To test this idea, synaptic transmission was measured during repetitive stimulation at moderate frequencies (14 Hz). During the initial phase of repetitive stimulation corresponding to post-stimulatory facilitation, the responses of mutant and wild-type mice were identical. However, after 10–15 stimuli, transmission at the mutant synapses became significantly depressed (Fig. 3), the maximum reduction being almost 50%. As the probability of neurotransmitter release at hippocampal synapses is below 0.5 per synaptic site^{12,13}, a given synapse will have released between 3 to 7 synaptic vesicles after about 15 stimuli, approximating the number of docked vesicles in resting

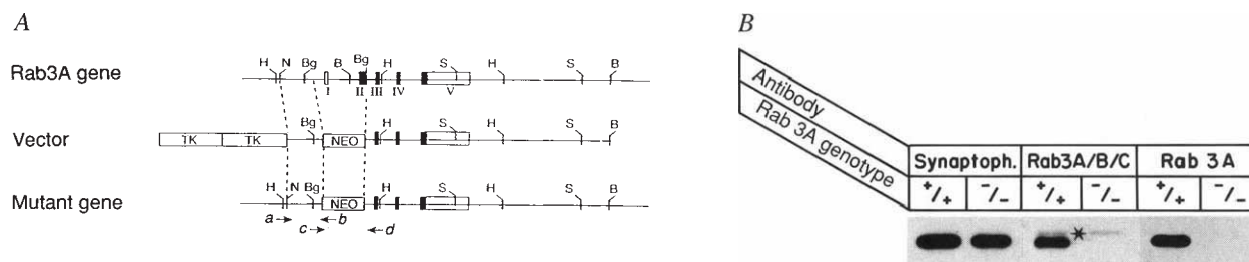


FIG. 1 Deletion of *rab3A* expression in mice by homologous recombination. **A**, Structures of the wild-type *rab3A* gene, the targeting vector (NEO, neomycin-resistance gene; TK, herpes simplex virus thymidine kinase gene), and the mutant *rab3A* gene after homologous recombination. Letters above the gene indicate positions of restriction enzyme sites (B, *Bam*HI; Bg, *Bgl*II; H, *Hind*III; N, *Nar*I; S, *Sac*I). Numbers below the gene identify exons, with filled boxes representing coding and open boxes non-coding sequences⁶. Arrows below the mutant gene labelled a to d identify positions of oligonucleotide primers used in PCRs selectively to amplify mutant (a and b) and wild-type *rab3A* alleles (c and d). **B**, Immunoblots of brain proteins from wild-type and homozygous mutant mice using an antibody against synaptophysin as a control (left lanes) and two different monoclonal antibodies against Rab3A, one of which (C142.1) reacts with Rab3A, B and C, whereas the second (C142.2) is specific for Rab3A¹⁸. Note that Rab3C (migrating above Rab3A; asterisk) is unchanged in mutant mice. **C**, Immunofluorescence localization of Rab3A (a, b), Rab3A/B/C (c, d), synaptophysin (e, f) and rabphilin (g, h) in frozen sections from the hippocampus of wild-type and Rab3A-deficient mice. Rab3A/B/C immunoreactivity in mutant mice lacking Rab3A exhibits a heterogeneous pattern with particularly high levels in the dentate gyrus. Rabphilin is drastically reduced in mutant mice (Fig. 4) apart from those regions showing high levels of RAB3B/C such as the dentate gyrus (h).

METHODS. The targeting vector for *rab3A* was constructed using murine *rab3A* genomic clones⁶ and standard methods^{8,22}. Mutant ES cell clones and chimaeric mice were obtained as described⁷⁻⁹. Wild-type and mutant alleles of *rab3A* were analysed by PCR using oligonucleotide primer pairs specific for the respective alleles. SDS-PAGE, immunoblotting and immunocytochemistry were done as described²²⁻²⁶. Scale bar, 1 mm.

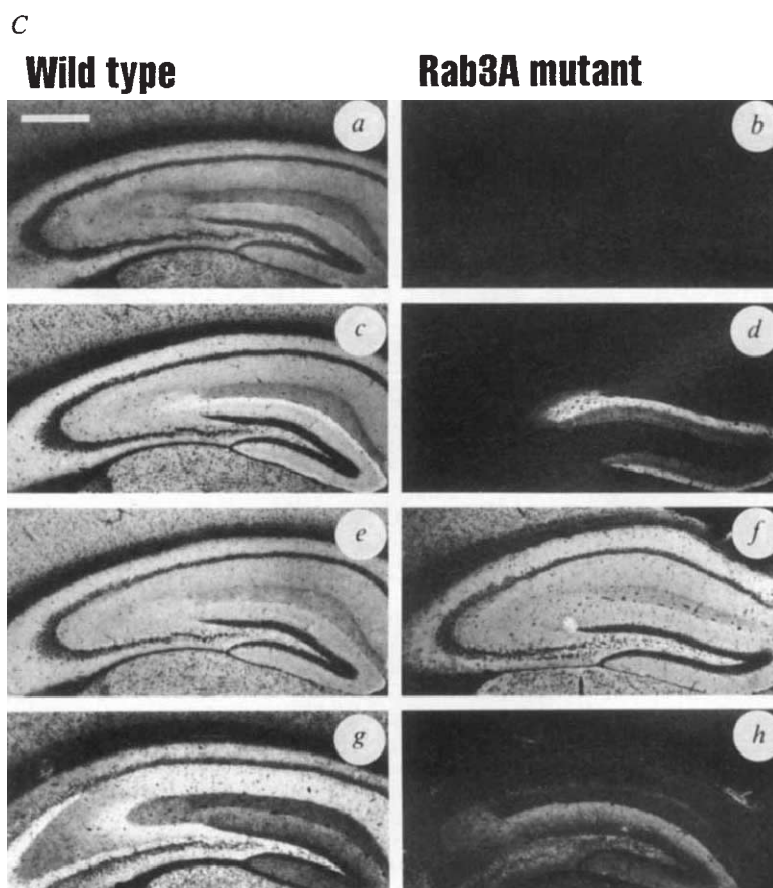


FIG. 2 Analysis of synaptic transmission and synaptic plasticity in *Rab3A*-deficient mice. *a*, E.p.s.c.s recorded in hippocampal slices from wild-type (wt; top) and mutant (mut; bottom) mice in response to paired stimuli delivered with different interstimulus intervals. *b*, Magnitude of paired-pulse facilitation as a function of interstimulus interval. Ratios of the second to the first e.p.s.c. amplitudes (mean $\pm$ s.e.m.) from wild-type (open circles; 4 animals, 16 cells) and mutant mice (filled circles; 8 animals, 34 cells) are plotted as a function of the interstimulus interval. *c*, Post-tetanic potentiation (PTP; mean $\pm$ s.e.m.) in wild-type (5 animals, 7 cells) and *rab3A* mutant mice (3 animals, 7 cells). *d*, LTP summary graphs (mean $\pm$ s.e.m.) for wild-type (open triangles; 5 animals, 10 cells) and mutant mice (filled triangles; 5 animals, 12 cells). LTP was induced by two 1-s trains of 100-Hz stimuli (arrow) applied 20 s apart. Neurons were kept current-clamped during tetanizations. There are no statistically significant differences in paired-pulse facilitation, PTP or LTP between wild-type and mutant mice.

METHODS. Hippocampal slices (150–200 μ m) prepared from 4–5-week-old mice were continuously superfused with artificial cerebrospinal fluid containing (in mM): 119 NaCl, 2.5 KCl, 1.0 MgSO_4 , 2.5 CaCl_2 , 1.0 NaH_2PO_4 , 26.2 NaHCO_3 , 10 glucose, 0.1 picrotoxin, adjusted to pH 7.3–7.4 in a 95% O_2 , 5% CO_2 atmosphere at 21–23 $^{\circ}\text{C}$. Whole-cell patch-clamp recordings^{27,28} were made from CA1 pyramidal cells (one cell per slice). Patch electrodes (5–10 $\text{M}\Omega$; no fire polishing) contained (in mM): 130 KCl, 5 NaCl, 1.1 EGTA, 1.0 MgCl_2 , 0.4 CaCl_2 , 10 HEPES-KOH, pH 7.2–7.3, 2 Mg-ATP , 0.1 Na-GTP . Drugs were applied by bath superfusion. To elicit e.p.s.c.s, afferent fibres in the stratum radiatum were stimulated at 0.1 Hz with bipolar tungsten electrodes. Holding potential was at -70 mV. Data acquisition and analysis were performed blind. In paired pulse experiments, 6–12 pairs of e.p.s.c.s were recorded and averaged at each interstimulus interval. The shortest interval was tested first. PTP was measured in the presence of AP5 (D-2-amino-5-phosphonopentanoic acid; 50 μM) by averaging the two largest e.p.s.c.s 1–2 s after a 1-s stimulation at 100 Hz, and dividing them by the average e.p.s.c. before the train. Summary LTP graphs were obtained by normalizing averaged e.p.s.c. amplitudes during 60-s intervals by a 10–15-min control baseline before LTP induction.

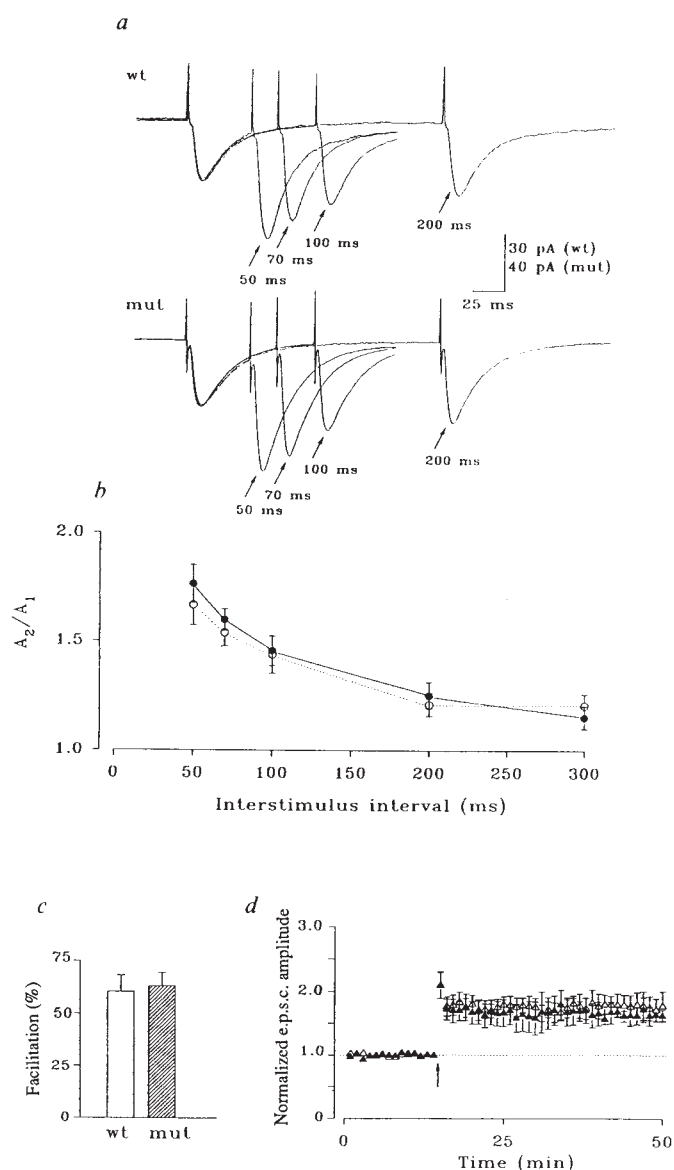
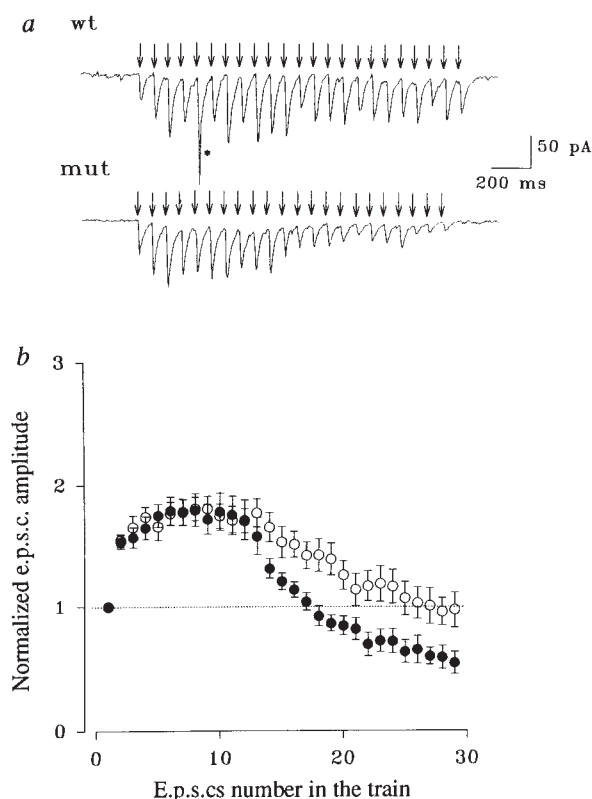


FIG. 3 Synaptic depression during repetitive stimulation in *rab3A*-deficient mice. *a*, e.p.s.c.s in response to repetitive stimuli delivered every 70 ms (14 Hz) recorded in slices from wild-type (top) and mutant mice (bottom). Asterisk in the top trace identifies an e.p.s.c. that triggered an action potential. *b*, Amplitudes of successive e.p.s.c.s plotted as a function of stimulus number during repetitive stimulation at 14 Hz from wild-type (open circles; 7 animals, 18 slices) and mutant (filled circles; 8 animals, 26 slices) mice. Normalized e.p.s.c. amplitudes are virtually identical for wild-type and mutant mice for the first 12 stimuli, after which mutant e.p.s.c.s become depressed. The differences between wild type and mutants become statistically significant ($P < 0.05$ t-test) beginning with the fourteenth e.p.s.c.

METHODS. Recordings were made as described for Fig. 2 in the presence of 50 μM AP5. Summary graphs were obtained by normalizing successive e.p.s.c. amplitudes to the amplitude of the first e.p.s.c. in the train. Three trains of e.p.s.c.s at the same 70-ms interstimulus interval were recorded in each neuron tested. Mean normalized values were then averaged for wild-type and mutant mice.

nerve terminals¹⁴. The Rab3A-deficient phenotype is thus apparent only after docked vesicles have been exhausted, suggesting a defect in synaptic vesicle docking. Although long-term potentiation (LTP) in the mutant mice was unchanged when tested with a strong induction protocol, weaker induction protocols could well reveal a defect in LTP.

Previous studies have indicated that Rab3A may have a function analogous to Sec4, which is essential for exocytosis in yeast^{1,15}. Although our results are consistent with a role for Rab3A in synaptic vesicle docking, it cannot be essential for exocytosis. Either its function must overlap with those of other synaptic vesicle proteins, or it must be limited to controlling the efficiency of release. To distinguish between these possibilities, we examined the expression of other Rab proteins in normal and mutant mice.

Four forms of Rab3 have been described: Rab3A and 3C are synaptic vesicle proteins^{5,16}, whereas Rab3B and 3D are primarily expressed outside the brain^{1,17}. A monoclonal antibody (C142.1) recognizing several forms of Rab3^{16,18} detects only small amounts of Rab3C in the mutant mice lacking Rab3A (Fig. 1B), and polyclonal antibodies showed only low levels of Rab3B and 3D. Rab3B, Rab3C and Rab3D were present in comparable amounts in wild-type and mutant mice (Fig. 1, and data not shown). The Rab3C detected by C142.1 in mutant mice is concentrated in selected synapses, for example those in the dentate gyrus (Fig. 1C, d). The immunoreactivity at most other regions is only faint, and many synapses give no signal.

To find out whether an unknown Rab protein might compensate for the loss of Rab3A, small GTP-binding proteins in synaptosomes from wild-type and mutant mice were quantified by blotting with GTP- γ^{35} S. Rab3A was found to be the major GTP-

binding Rab protein in synaptosomes, accounting for at least 25% of the total Rab protein GTP-binding capacity (Fig. 4a). No significant amount of GTP-binding protein of the same molecular weight was detected in mutant mice. Other GTP-binding protein signals were similar in wild-type and mutant mice, indicating that their levels are unaffected by the *rab3A* deletion.

Rab3A may act on rabphilin, a Ca^{2+} -binding protein that binds to Rab3A in a GTP-dependent manner¹⁹. Immunoblotting shows that rabphilin levels in mutant mice are $30.4 \pm 2.8\%$ ($n = 6$) of those of wild-type mice (Fig. 4b). A similar reduction was observed by immunocytochemistry (Fig. 1C), but brain regions with large amounts of Rab3C, such as the dentate gyrus, had almost normal levels of rabphilin, indicating that the decrease elsewhere may reflect the absence of Rab3A. The mutant mice had normal levels of more than 20 other synaptic proteins, including Rab3A-binding proteins (GDI (GDP-dissociation inhibitor) and MSS4 (mammalian suppressor of *sec4*))¹⁻³, other GTP-binding proteins (Rab1, Rab5, dynamin)¹⁻³, and proteins whose yeast homologues interact with the yeast Rab protein Sec4 (Munc18-1, syntaxin, SNAP-25)^{20,21}. RNA blotting showed that the abundance of rabphilin messenger RNA was unchanged in the mutant mice (data not shown). Although rabphilin translation may be selectively decreased in the mutant mice, it seems more likely that rabphilin is less stable in these animals.

To determine whether the remaining rabphilin in the mutant mice is associated with a compensating GTP-binding protein, rabphilin was immunoprecipitated with and without GTP- γ S from forebrain homogenates (Fig. 4d). GTP-dependent co-precipitation of Rab3A was nearly stoichiometric in wild-type mice. In the mutants, however, no co-precipitated GTP-binding pro-

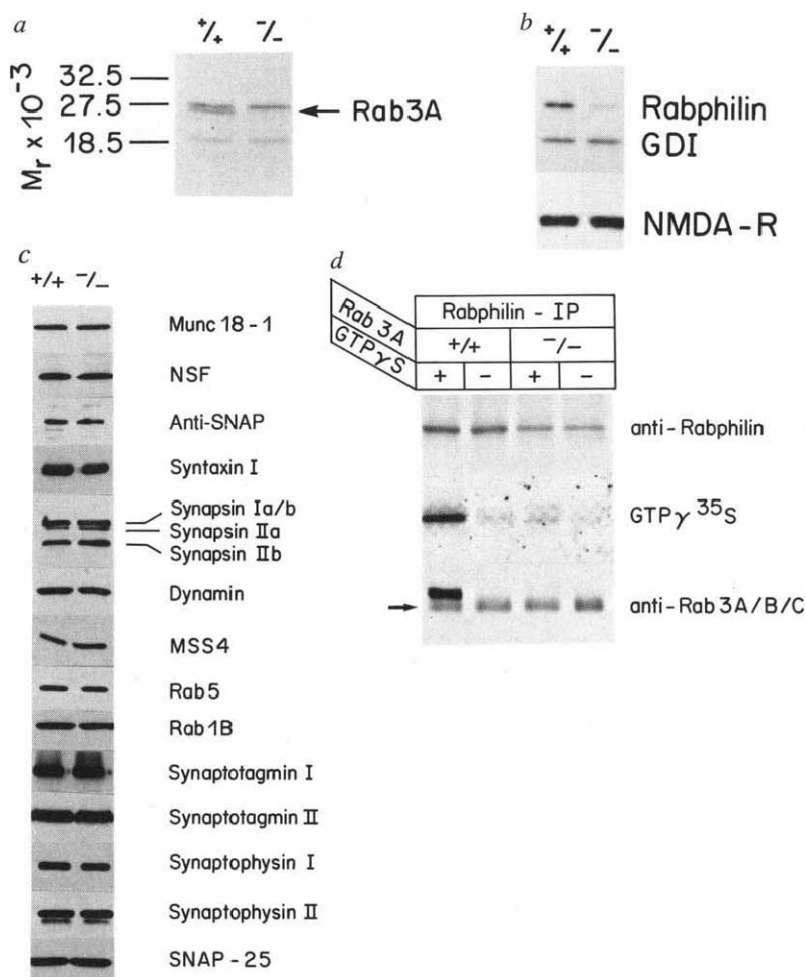


FIG. 4 Selective decrease in rabphilin in *rab3A*-deficient mice. **a**, Analysis of synaptosomal GTP-binding proteins from wild-type (+/+) and mutant mice (-/-) by blotting with GTP- γ^{35} S (used instead of 32 P-labelled GTP for better resolution) and autoradiography. Arrow points to the $M_r = 25,000$ band that corresponds to Rab3A and is selectively deleted in the mutant mice. Phosphorimager quantification showed that Rab3A accounts for 23–27% of the total Rab-GTP binding, and that the other components do not change after Rab3A deletion (data not shown). Longer exposures did not reveal any other Rab protein with the same electrophoretic mobility as Rab3A in the mutant mice. **b** and **c**, Immunoblot of brain proteins from wild-type and Rab3A mutant mice with a panel of antibodies raised against synaptic proteins demonstrates a selective decrease in rabphilin in the mutant mice to 30% of wild-type levels (normalized for GDI as control). Note that other Rab3-interacting proteins (GDI, MSS4) and synaptic proteins are unchanged. **d**, GTP- γ S-dependent coprecipitation of Rab3A with rabphilin in wild-type but not mutant mice. Immunoprecipitates (IP) with rabphilin antibodies were analysed by immunoblotting with antibodies against rabphilin (top) and Rab3A/B/C (C142.1, bottom; the faint band in all lanes (arrow) corresponds to immunoglobulin light chains) and by blotting with GTP- γ^{35} S to reveal any other coprecipitating Rab proteins (middle). In *rab3A*-deficient mice, longer exposure did not detect any other GTP-binding protein coprecipitating with rabphilin (data not shown).

METHODS. Total brain proteins from mutant and wild-type mice were analysed by SDS-PAGE and blotting with GTP- γ S and a panel of antibodies^{6,9,13,15,26,29,30}. Proteins were quantified using iodinated secondary antibodies and phosphorimager detection; rabphilin was estimated in samples from six pairs of wild-type and mutant mice and normalized for GDI on the same blot. Rabphilin was immunoprecipitated as described³⁰ and analysed by SDS-PAGE and immunoblotting or blotting with GTP- γ^{35} S.

tein was detectable by blotting with GTP- $\gamma^{35}\text{S}$ or with anti-Rab3A/B/C antibody (Fig. 4d). This result confirms that Rab3A is the major target of rabphilin but does not exclude binding of Rab3C to rabphilin because of the low abundance of Rab3C in forebrain.

Mice lacking Rab3A thus show a 70% decrease in rabphilin levels (probably due to enhanced degradation) and premature depression of synaptic transmission during short trains of repetitive stimuli. These changes are selective because the levels of more than 20 other synaptic proteins are normal, as are all other synaptic parameters tested, and synaptic structure and connectivity appear to be unaffected. Other Rab proteins do not seem to compensate for the absence of Rab3A. We cannot rule out the possibility that another Rab protein shares at least one of Rab3A's functions, but such a protein would have to be expressed at a uniform, low level in all synapses. Rab3A is thus dispensable for synaptic vesicle exocytosis but is important for regulating synaptic efficiency, probably by interaction with rabphilin. The pattern of synaptic depression observed in the mutant mice indicates that Rab3A may assist in the formation of a pre-fusion complex between the synaptic vesicle and the plasma-membrane. □

Received 5 April; accepted 28 April 1994.

1. Takei, Y., Kaibuchi, K., Kikuchi, A. & Kawata, M. *Int. Rev. Cytol.* **133**, 187–230 (1992).
2. Hall, A. *Science* **249**, 635–640 (1990).
3. Simons, K. & Zerial, M. *Neuron* **11**, 789–799 (1993).
4. Fischer von Mollard, G. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1988–1992 (1990).
5. Fischer von Mollard, G., Südhof, T. C. & Jahn, R. *Nature* **349**, 79–81 (1991).

6. Baumert, M., Fischer von Mollard, G., Jahn, R. & Südhof, T. C. *Biochem. J.* **293**, 157–163 (1993).
7. Robertson, E. J. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 71–112 (IRL, Oxford, 1987).
8. Soriano, P., Montgomery, C., Geske, R. & Bradley, A. *Cell* **64**, 693–702 (1991).
9. Rosahl, T. W. et al. *Cell* **75**, 661–670 (1993).
10. Bliss, T. V. P. & Collingridge, G. L. *Nature* **361**, 31–39 (1993).
11. Hawkins, R. D., Kandel, E. R. & Siegelbaum, S. A. *Rev. Neurosci.* **16**, 625–665 (1993).
12. Rosenmund, C., Clements, J. D. & Westbrook, G. L. *Science* **262**, 754–757 (1993).
13. Hessler, N. A., Shirkar, A. M. & Malinow, R. *Nature* **366**, 569–572 (1993).
14. Peters, A., Palay, S. L. & Webster, H. deF. *The Fine Structure of the Nervous System* 3rd edn (Oxford Univ. Press, Oxford and New York, 1991).
15. Südhof, T. C. & Jahn, R. *Neuron* **6**, 665–677 (1991).
16. Fischer von Mollard, G., Stahl, B., Khokhachev, A., Südhof, T. C. & Jahn, R. *J. biol. Chem.* **269**, 10971–10974 (1994).
17. Baldini, G., Hohl, T., Lin, H. Y. & Lodish, H. F. *Proc. natn. Sci. U.S.A.* **89**, 5049–5052 (1992).
18. Matteoli, M. et al. *J. Cell Biol.* **115**, 625–633 (1991).
19. Shirataki, H. et al. *J. biol. Chem.* **267**, 10946–10949 (1992).
20. Soellner, T. et al. *Nature* **362**, 318–324 (1993).
21. O'Connor, V., Augustine, G. J. & Betz, H. *Cell* **76**, 785–787 (1994).
22. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
23. Mignery, G. A., Südhof, T. C., Takei, K. & De Camilli, P. *Nature* **342**, 192–195 (1989).
24. Johnston, P. A., Jahn, R. & Südhof, T. C. *J. biol. Chem.* **264**, 1268–1273 (1989).
25. De Camilli, P., Cameron, R. & Greengard, P. *J. Cell Biol.* **96**, 1337–1354 (1983).
26. Takei, K. et al. *J. Neurosci.* **12**, 489–505 (1992).
27. Blanton, M. G., Lo Turco, J. J. & Kriegstein, A. R. *J. Neurosci. Meth.* **30**, 203–210 (1989).
28. Bolshakov, V. Y. & Siegelbaum, S. A. *Science* **264**, 1148–1152 (1994).
29. Burton, J., Roberts, D., Montaldi, M., Novick, P. & De Camilli, P. *Nature* **361**, 464–467 (1993).
30. Fykse, E. M. et al. *J. Neurosci.* **13**, 4997–5007 (1993).

ACKNOWLEDGEMENTS. We thank M. S. Brown, J. L. Goldstein, E. Kandel, B. Armitage, J. Burton and J. Herz for comments and advice; K. Kirk, H. Lodish, N. Brose, R. Jahn, E. Fykse, Y. Hata, W. Balch and C. Li for antibodies; and L. Lundquist, A. Roth, I. Leznicki and L. Daniels for technical assistance. This study was supported by grants from the W. M. Keck Foundation, the Perot Family Foundation (to T.C.S. and R.E.H.), and the US Public Health Service (to P.D.C.). M.G. was supported by a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft.

NF-AT components define a family of transcription factors targeted in T-cell activation

Jeffrey P. Northrop, Steffan N. Ho, Lel Chen, Daryl J. Thomas, Luika A. Timmerman, Garry P. Nolan*, Arle Admon†‡ & Gerald R. Crabtree

Departments of Developmental Biology and Pathology, Howard Hughes Medical Institute, and * Department of Molecular Pharmacology, Stanford University Medical School, Stanford, California 94305, USA

† Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720, USA

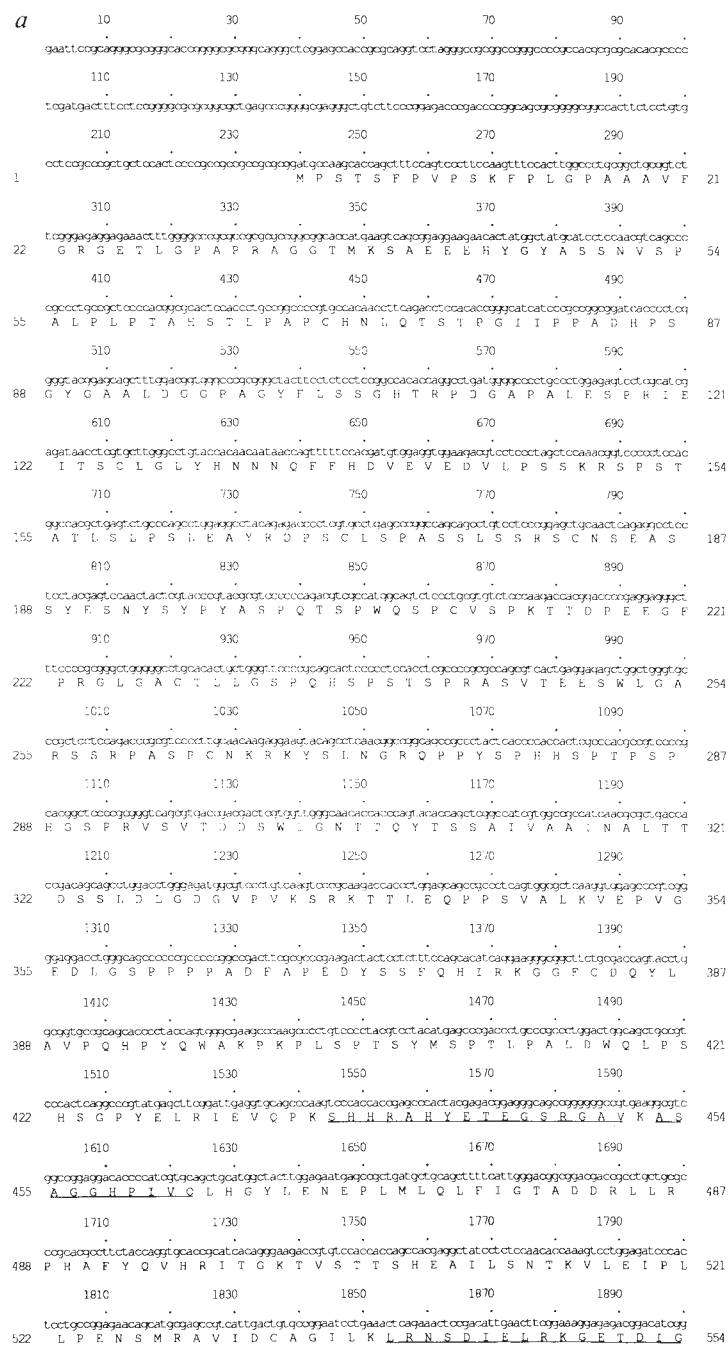
THE NF-AT transcription complex¹ is required for the expression of a group of proteins that collectively coordinate the immune response^{2,4}. Here we purify two proteins encoded by separate genes that represent the pre-existing (p) and cytosolic (c) components of NF-AT. Expression of the full-length complementary DNA encoding NF-AT_c activates the interleukin (IL-2) promoter in non-T lymphocytes, whereas a dominant negative of NF-AT_c specifically blocks activation of the IL-2 promoter in T lymphocytes, indicating that NF-AT_c is required for IL-2 gene expression. NF-AT_c RNA expression is largely restricted to lymphoid tissues and is induced upon T-cell activation. The other protein, NF-AT_p, is highly homologous to NF-AT_c over a limited domain which shows similarity to the Dorsal/Rel family⁵, but has a wider tissue distribution. Agents that increase intracellular Ca²⁺ or activate protein kinase C independently modify NF-AT_c, indicating that distinct signalling pathways converge on NF-AT_c to regulate its function.

As the cytosolic component of NF-AT is present at low concentrations in human lymphoid cell lines⁶, we purified NF-AT_c

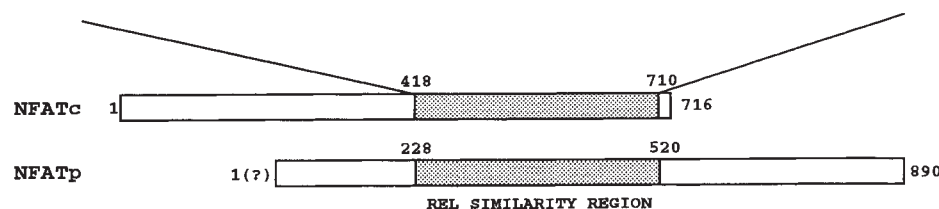
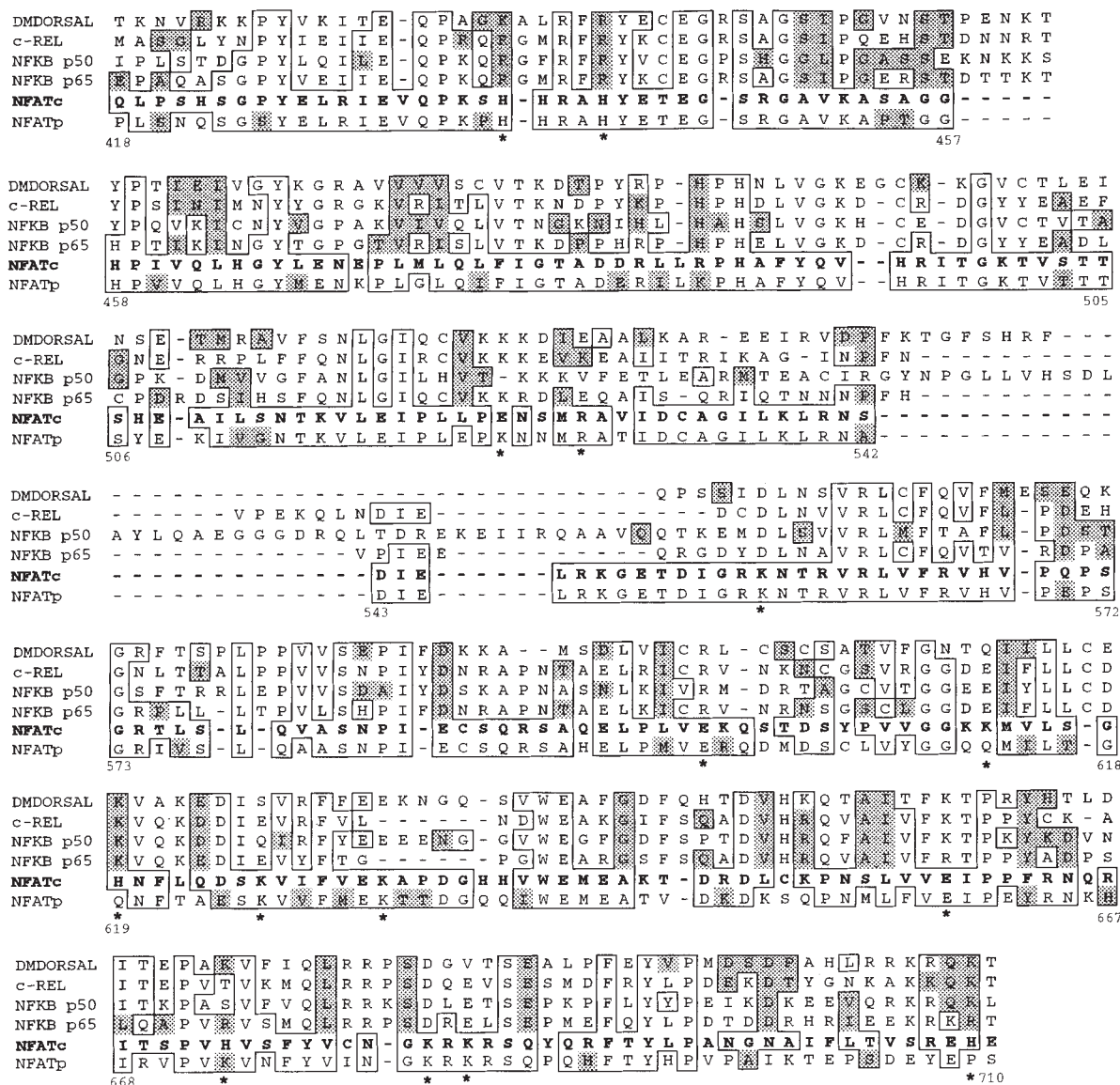
from bovine thymus. Amino-acid sequence from six peptides was used to isolate human cDNA clones spanning 2,742 nucleotides (Fig. 1a). The cDNA encodes a protein of 716 amino acids with a predicted relative molecular mass (M_r) of 77,870. An upstream in-frame stop codon indicates that the entire NF-AT_c protein is encoded by this cDNA. A unique repeated sequence of 13 residues (241–253 and 290–302) was identified. The carboxy-terminal half of NF-AT_c shows ~20% identity with the DNA-binding and dimerization regions of the Dorsal/Rel family of transcription factors (Fig. 1b; see ref. 5 for review). There are a significant number of charge reversals between the Rel family members and NF-AT_c, suggesting that charge might be conserved at these positions to maintain specific protein or DNA interactions. Six additional peptides from our purified bovine protein (data not shown) are derived from the bovine homologue of NF-AT_p. (A murine cDNA fragment of NF-AT_p has been reported since submission of this letter⁷.) Comparison of NF-AT_c and NF-AT_p reveals that they are products of different genes with 73% identity in the Rel-similarity region (Fig. 1b), but there is little similarity outside this region. A murine cDNA for NF-AT_c was isolated and the predicted protein was found to be 87% identical to human NF-AT_c (data not shown), and distinctly different from murine NF-AT_p.

NF-AT_c messenger RNA is absent in HeLa cells (Fig. 2a, lane 7), a cell line incapable of IL-2 or NF-AT-dependent transcription, but is inducible in Jurkat cells (Fig. 2a). This induction is sensitive to cyclosporin A (CsA), indicating that NF-AT_c may participate in an auto-stimulatory loop because CsA blocks its nuclear association¹. A variety of non-T cell lines do not express NF-AT_c mRNA (Fig. 2b), consistent with the observed T-cell-restricted pattern of IL-2 transcription and NF-AT activity^{2,8–10}. NF-AT-dependent transcription predominates in spleen, thymus and skin of transgenic mice expressing an NF-AT-dependent transgene⁸. Consistent with this, murine NF-AT_c mRNA shows the same pattern of expression (Fig. 2c, top panel). Low levels of NF-AT_c expression are seen in lung and heart, but this may be due to contamination with circulating T cells. Murine NF-AT_p mRNA was comparably expressed in brain, heart, thymus and

‡ Present address: Department of Biology, Technion Institute, Haifa 32000, Israel.



b



spleen, but in contrast to NF-AT_c, NF-AT_p was not inducible by phorbol myristyl acetate (PMA) and ionomycin (Fig. 2c, bottom panel).

Although the predicted size of NF-AT_c is less than the observed size of $M_r \sim 100$ – 120 K (ref. 6), transfected NF-AT_c migrates at ~ 105 K, confirming that the cDNA can encode a protein with an apparent size similar to the native protein (Fig. 2d).

NF-AT_c expressed in non-T-cell lines specifically activated transcription from the NF-AT site and the IL-2 promoter (Fig. 3a, left, and b). In Jurkat cells, overexpression of NF-AT_c activated an NF-AT-dependent promoter but not an HNF-1-dependent promoter (Fig. 3a, right) or an AP-1-dependent promoter (data not shown). Transfection of the NF-AT_c cDNA

gives rise to DNA-binding activity that is indistinguishable from endogenous NF-AT (Fig. 3c, lanes 1–4). Antibody directed against the encoded epitope induces a supershift of the NF-AT complex (data not shown), indicating that NF-AT_c participates in this activity. Nuclear NF-AT activity in transfected COS cells comigrates with, and has the same binding specificity as, the native NF-AT complex in T cells (Fig. 3c, lanes 4–11). NF-AT_c from transfected COS cells reconstitutes NF-AT DNA-binding activity (Fig. 3c, lanes 12–16), as does NF-AT_c from T cells^{1,6}. A monoclonal antibody (7A6) raised against bacterially expressed NF-AT_c induced a supershift of the majority of the endogenous NF-AT gel-shift complex from murine thymus (Fig. 3d) and spleen (data not shown). The inability of the antibody to induce a quantitative supershift suggests that the NF-AT complex is

heterogeneous. As the antibody epitope is outside the region of homology between NF-AT_c and NF-AT_p (data not shown), the NF-AT complex not supershifted may, in part, represent NF-AT_p. These data are consistent with the identification of peptides at approximately equimolar ratios from both NF-AT_c and NF-AT_p in the purified protein from bovine thymus.

A dominant negative NF-AT_c consisting of amino acids 1–418 was identified by deletion analysis of the cDNA (S.N.H. *et al.*, manuscript in preparation). This region of the cDNA, which is not found in NF-AT_p, was used to assess the contribution of NF-AT_c to the activation of the IL-2 gene. Expression of the dominant negative resulted in more than 90% inhibition of IL-2 promoter function, as well as transcription directed by the NF-AT site (Fig. 4a). This effect was highly specific, because transcription directed by the AP-1 site or the Rous sarcoma virus promoter and enhancer was unaffected (Fig. 4a). These results indicate that NF-AT_c contributes substantially to IL-2 gene expression in T cells.

Post-translational modification of NF-AT_c was investigated in cells treated with agents that activate protein kinase C or increase intracellular Ca²⁺. The bulk of NF-AT_c in cells treated

with ionomycin migrates faster than that in non-treated cells and this mobility shift is inhibited by CsA (Fig. 4), lanes 1, 3, 4). This is consistent with a dephosphorylation event, possibly resulting from direct action by calcineurin¹¹, but any one of a large number of processes could produce the observed mobility changes. There is preliminary evidence that NF-AT_p is a substrate for calcineurin, but the mobility shifts produced by phosphatase treatment of NF-AT_p (ref. 12) or NF-AT_c (L.C. and G.R.C. unpublished results) are far greater than those shown in Fig. 4. These observations suggest that NF-AT_c may not be a direct substrate of calcineurin. PMA treatment produces a slower-migrating NF-AT_c (Fig. 4b, lane 2), so pathways promoted by protein kinase C may contribute to NF-AT activity by modification of NF-AT_c as well as by activation of the nuclear component.

The full-length clone for NF-AT_c and the cDNA fragment for NF-AT_p (ref. 7) seem to define a family of transcription factors distantly related to Dorsal/Rel proteins. Features of NF-AT_c common to the Dorsal/Rel group include cytoplasmic-to-nuclear translocation, responsiveness to environmental stimuli, and the presence of certain signature sequences within the DNA

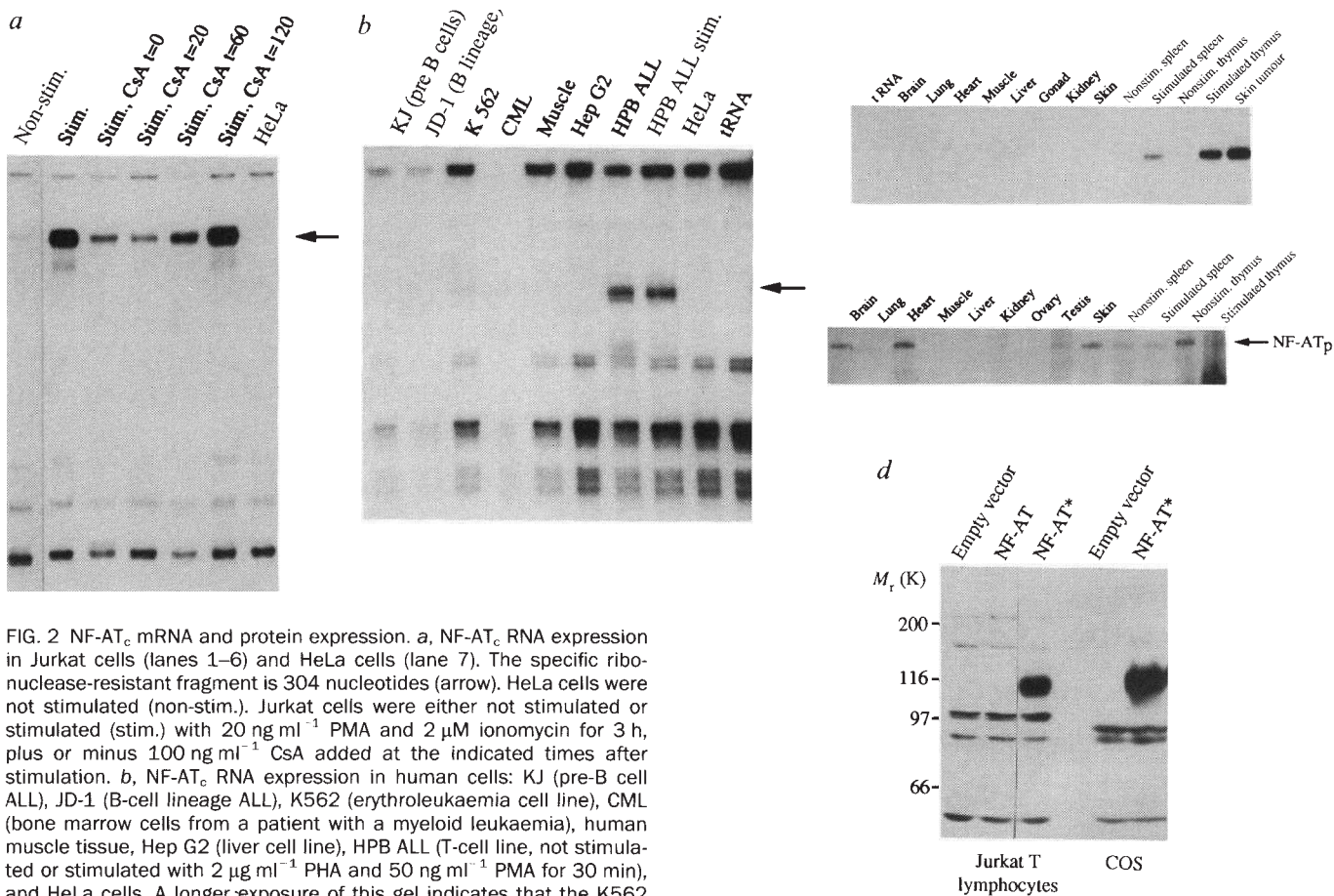
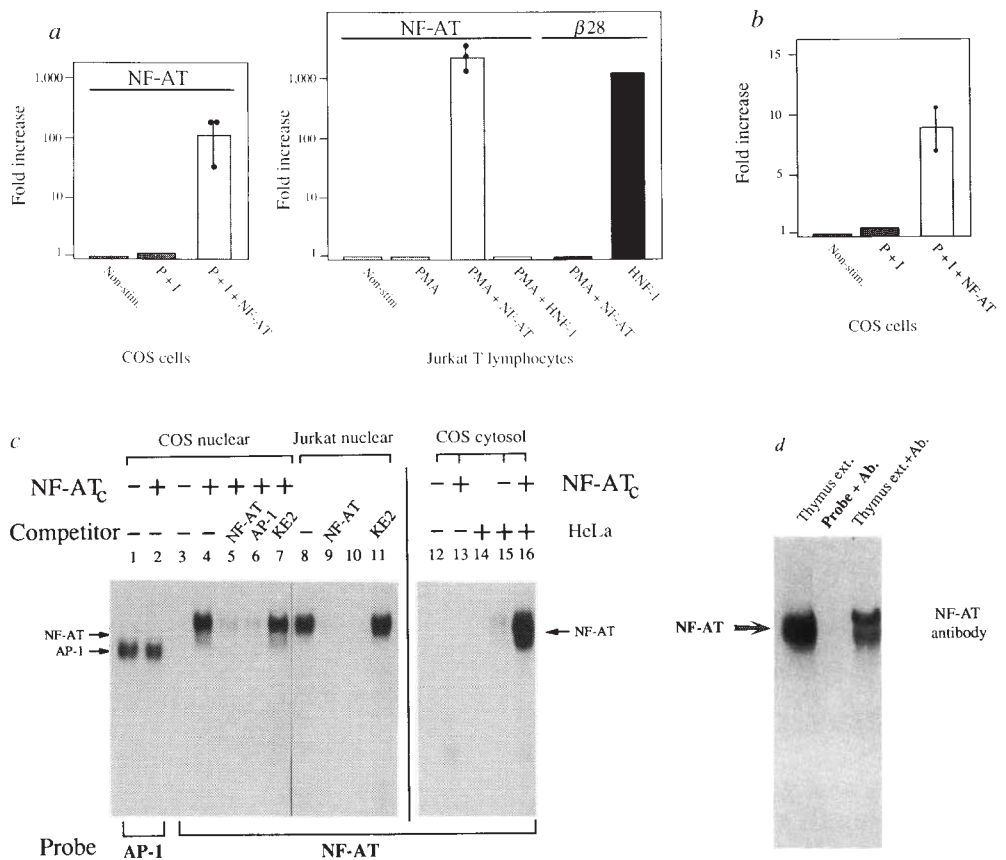


FIG. 2 NF-AT_c mRNA and protein expression. *a*, NF-AT_c RNA expression in Jurkat cells (lanes 1–6) and HeLa cells (lane 7). The specific ribonuclease-resistant fragment is 304 nucleotides (arrow). HeLa cells were not stimulated (non-stim.). Jurkat cells were either not stimulated or stimulated (stim.) with 20 ng ml⁻¹ PMA and 2 μM ionomycin for 3 h, plus or minus 100 ng ml⁻¹ CsA added at the indicated times after stimulation. *b*, NF-AT_c RNA expression in human cells: KJ (pre-B cell ALL), JD-1 (B-cell lineage ALL), K562 (erythroleukaemia cell line), CML (bone marrow cells from a patient with a myeloid leukaemia), human muscle tissue, Hep G2 (liver cell line), HPB ALL (T-cell line, not stimulated or stimulated with 2 μg ml⁻¹ PHA and 50 ng ml⁻¹ PMA for 30 min), and HeLa cells. A longer exposure of this gel indicates that the K562 cell line contains a small amount of NF-AT_c RNA (data not shown). *c*, NF-AT_c (top) and NF-AT_p (bottom) RNA expression in mouse tissues and a skin tumour derived from NF-AT-Tag transgenic mice⁸. Cells were either not stimulated or stimulated with 20 ng ml⁻¹ PMA and 2 μM ionomycin for 3 h. The predicted size of the fragment homologous to the probe is indicated. *d*, Expression of haemagglutinin (HA) epitope-tagged NF-AT_c. Jurkat Tag cells⁶ or COS cells were transiently transfected with pBJ5 expression vector alone (empty vector), expression vector for non-epitope-tagged NF-AT_c (NF-AT) or epitope-tagged NF-AT_c (NF-AT*). METHODS. *a*–*c*, Specific human or mouse NF-AT_c cDNA fragments or mouse NF-AT_p cDNA fragments derived by PCR were used as templates

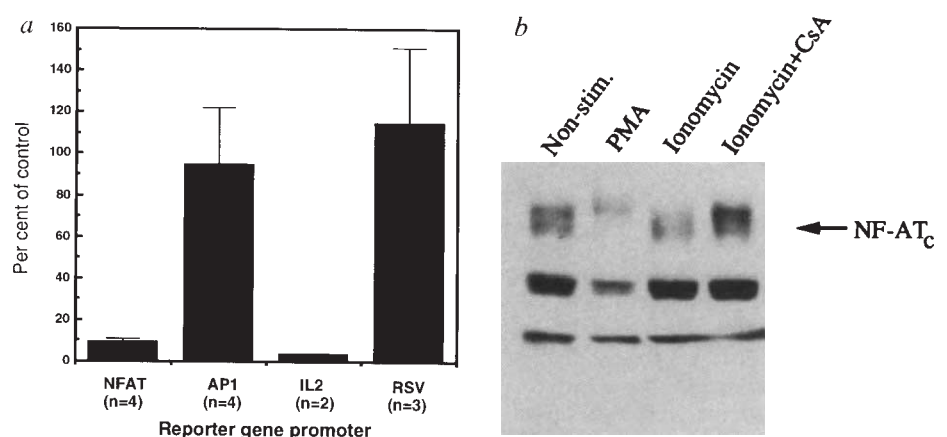
for synthesis of RNA transcripts. Ribonuclease protection was according to ref. 17 using 10 μg total RNA. Splenocytes and thymocytes were isolated and treated as described⁸ before isolating RNA, otherwise whole tissue was used. *d*, Jurkat Tag cells⁶ or COS cells were transfected with 3 μg of each expression vector as described⁶. Cells lysates were analysed for HA-tagged NF-AT_c by western blotting with the anti-HA antibody 12CA5. An expression vector for NF-AT_c was constructed by inserting the entire cDNA into the vector pB15 (ref. 18). An expression vector for epitope-tagged NF-AT_c was constructed by replacing the first 11 amino acids of NF-AT_c with the sequence MFYPYDVPDYA MSEPKAIDPKLS (HA epitope is underlined) in the vector pBJ5.

FIG. 3 Functional expression of NF-AT_c. *a* and *b*, COS and Jurkat cells transfected with reporter constructs for NF-AT, HNF-1 (β 28) or IL-2, and expression vectors for NF-AT_c (+NF-AT), HNF-1 α (+HNF-1) or empty pBJ5 vector. Cells were stimulated as indicated (P+I, PMA plus ionomycin). Data are expressed as fold induction of luciferase activity over the non-stimulated value with empty pBJ5 vector. Bars represent mean and range of 2–3 independent transfections. *c*, NF-AT_c expression in COS cells gives rise to specific DNA-binding activity. Gel mobility shifts using nuclear extracts from COS cells transfected with pBJ5 (lanes 1 and 3), with NF-AT_c (lanes 2 and 4–7), from non-transfected Jurkat cells (lanes 8–11), or using cytosols from pBJ5- or NF-AT_c-transfected COS cells (lanes 12, 13, 15, 16) combined with HeLa nuclear extract (lanes 15, 16). Lane 14, HeLa nuclear extract alone. Labelled AP-1 (lanes 1, 2) or NF-AT (lanes 3–16) probes and cold competitor oligonucleotides are indicated. Arrows indicate specific AP-1 and NF-AT complexes. *d*, Antibody-induced super-shift of NF-AT. The monoclonal antibody 7A6 (IgG1) was used in gel mobility shift assays with extracts (ext.) from stimulated murine thymocytes. The NF-AT complex and the antibody–NF-AT complex are indicated by arrows. The antibody did not interact with probe alone (lane 2) and did not alter either an AP1 or an NF- κ B gel-shift complex (data not shown). Ab, antibody.



described^{6,19}. *c* and *d*, COS cells were transfected with epitope-tagged NF-AT_c as described for Fig. 2. COS cells, Jurkat cells and murine thymocytes were stimulated for 3 h with PMA and ionomycin; HeLa cells were stimulated for 3 h with PMA alone and nuclear extracts prepared⁹. Cytosol was prepared from non-stimulated COS cells. Gel mobility shifts were determined as described^{1,6}. Monoclonal antibody 7A6 was derived from a mouse immunized with a bacterially expressed glutathione S-transferase fusion protein containing NF-AT_c residues 1–654.

FIG. 4 *a*, NF-AT_c dominant negative mutant assayed by transient transfection. Jurkat Tag cells were transfected with vector plasmid (control) or with the dominant negative NF-AT_c plasmid plus the indicated secreted alkaline phosphatase reporter plasmid. Transfected cells were transferred to fresh culture medium 24 h after transfection and secreted alkaline phosphatase activity was measured¹¹ 16–24 h later after stimulation with 1 μ M ionomycin plus 20 ng ml⁻¹ PMA (NF-AT and IL-2 reporters), 20 ng ml⁻¹ PMA alone (AP-1 reporter) or no stimulation (Rous sarcoma virus (RSV) reporter). Bars indicate secreted alkaline phosphatase activity from cells transfected with dominant negative NF-AT_c as a percentage of the activity from cells transfected in parallel with control plasmid, and represent data obtained from independent transfections (*n*). The dominant negative NF-AT_c consists of a C-terminal truncation of the epitope-tagged NF-AT_c expression plasmid (Fig. 2d) extending to the *Pvu*II site at amino acid 418. *b*, Changes in mobility of epitope-tagged NF-AT_c expressed in Jurkat cells. Cells



were transfected with NF-AT_c as for Fig. 2d and stimulated with 20 ng ml⁻¹ PMA or 2 μ M ionomycin $\pm$ 100 ng ml⁻¹ CsA for 2 h. Whole-cell lysates were analysed by western blotting as for Fig. 2d.

binding and dimerization domain of these proteins. The roles of NF-AT_c and NF-AT_p are as yet not defined, but they could participate in sequential events of thymic maturation^{13,14} or could direct distinct patterns of cytokine expression in mature cells through differential expression in various T-cell subsets or by alternative pairing through the Rel-like domain. In many mature cells, the partner (NF-AT_n) appears to be AP-1 (refs 6, 15), but, these studies do not rule out the existence of mediators of NF-AT_n activity other than AP-1. As both protein kinase C- and calcium-induced signalling pathways appear to modify NF-AT_c independently, the function of NF-AT_c might be to integrate signals derived from different cell membrane receptors. □

Received 1 October 1993; accepted 8 April 1994.

1. Flanagan, W. F., Cortes, B., Bram, R. J. & Crabtree, G. R. *Nature* **352**, 803–807 (1991).
2. Shaw, J.-P. et al. *Science* **241**, 202–205 (1988).
3. Crabtree, G. R. *Science* **243**, 355–361 (1989).
4. Weiss, A. *Cell* **73**, 209–212 (1993).
5. Nolan, G. P. & Baltimore, D. *Curr. Biol.* **2**, 211–220 (1992).
6. Northrop, J. P., Ullman, K. S. & Crabtree, G. R. *J. biol. Chem.* **268**, 2917–2923 (1993).
7. McCaffrey, P. G. et al. *Science* **262**, 750–754 (1993).
8. Verweij, C. L., Guidos, C. & Crabtree, G. R. *J. biol. Chem.* **265**, 15788–15795 (1990).
9. Flier, S. et al. *Genes Dev.* **4**, 1823–1834 (1990).
10. Yaseen, N. R., Maizel, A. L., Wang, F. & Sharma, S. *J. biol. Chem.* **268**, 14285–14293 (1993).
11. Clipstone, N. A. & Crabtree, G. T. *Nature* **357**, 695–697 (1992).
12. Jain, J. et al. *Nature* **365**, 352–355 (1993).
13. Chen, D. & Rothenberg, E. V. *Molec. cell. Biol.* **13**, 228–237 (1993).
14. Riegel, J. S., Richie, E. R. & Allison, J. P. *J. Immun.* **144**, 3611–3618 (1990).
15. Jain, J., McCaffrey, P. G., Valge-Archer, V. E. & Rao, A. *Nature* **356**, 801–804 (1992).
16. Miyata, T., Mizunawa, S. & Yasunaga, T. *J. molec. Evol.* **12**, 214–236 (1979).
17. Melton, D. A. *Nucleic Acids Res.* **12**, 7035–7056 (1984).
18. Mendel, D. B., Hansen, L. P., Graves, M. K., Conley, P. B. & Crabtree, G. R. *Genes Dev.* **5**, 1042–1056 (1991).
19. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044–1050 (1982).
20. de Wet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R. & Subramani, S. *Molec. cell. Biol.* **7**, 725–737 (1987).

ACKNOWLEDGEMENTS. J.P.N. and S.N.H. contributed equally to this work. We thank M. Flanagan for assistance with purification, P. Kao and B. Cortes for mutant NF-AT affinity resin; D. Spencer, P. Khavari and N. Clipstone for discussion; L. Hansen for pBJ5 HNF-1 α ; M. Cleary and H. Blau for human RNA samples; A. Krensky for the human PBL library; and S. LaSavia for help with obtaining calf thymuses. This work was supported by the Howard Hughes Medical Institute (J.P.N., L.C., D.J.T., L.A.T., G.R.C.), the NIH (G.R.C.), and the Walter V. and Idun Y. Berry Fellowship (S.N.H.).

SH2 domain specificity and activity modified by a single residue

Luc E. M. Marengere^{*†}, Zhou Songyang[†],
Gerald D. Gish^{*}, Michael D. Schaller[§],
J. Thomas Parsons[§], Michael J. Stern^{||},
Lewis C. Cantley[†] & Tony Pawson^{*†¶}

^{*} Division of Molecular and Developmental Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

[†] Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada

[‡] Department of Cell Biology, Harvard Medical School and Department of Medicine, Beth Israel Hospital, Boston, Massachusetts 02115, USA

[§] Department of Microbiology and Cancer Center, University of Virginia, Charlottesville, Virginia 22908, USA

^{||} Boyer Center for Molecular Medicine, Yale University, 295 Congress Avenue, PO Box 9812, New Haven, Connecticut 06536-0812, USA

MANY intracellular targets of protein-tyrosine kinases possess Src homology 2 (SH2) domains that directly recognize phosphotyrosine-containing sites on autophosphorylated growth factor receptors and cytoplasmic proteins, and thereby mediate the activation of biochemical signalling pathways^{1–7}. SH2 domains possess relatively well conserved residues that form the phosphotyrosine-binding pocket^{8–11}, and more variable residues that are implicated in determining binding specificity by recognition of the three amino acids carboxy-terminal to phosphotyrosine (the +1 to +3 positions)^{5,7,12,13}. One such residue, occupying the EF1 position of the +3-binding pocket, is a Thr in the SH2 domain of the Src tyrosine kinase¹², but is predicted to be a Trp in the SH2 domain of the Sem-5/drk/Grb2 adaptor protein⁵. Here we report that changing this residue in the Src SH2 domain from Thr to Trp switches its selectivity to resemble that of the Sem-5/drk/Grb2 SH2 domain. Furthermore, this mutant Src SH2 domain effectively substitutes for the SH2 domain of the Sem-5 protein in activation of the Ras pathway *in vivo*. These results identify a residue that can modify SH2 selectivity, and indicate that the biological activity of an SH2 domain correlates with its binding specificity.

The wild-type (wt) Src SH2 domain preferentially selects the sequence pYEEI (Glu at the +1 position, Glu at +2 and Ile at +3) from a degenerate phosphopeptide library⁵. ThrEF1 of the Src SH2 domain is located on the edge of a hydrophobic pocket that binds the +3Ile and contacts the +3Ile side chain¹² (Fig. 1a). In the SH2 domain of the *Caenorhabditis elegans* protein Sem-5¹⁴, and its *Drosophila* and mammalian homologues (drk^{15,16}, Grb2¹⁷), the predicted EF1 residue is a Trp (Fig. 1b). In contrast to Src, the Sem-5 SH2 domain preferentially selects pY-(L/V)-N-(V/P), with the +2Asn dominating the selection⁵. drk and Grb2 select similar SH2-binding sequences to Sem-5¹⁸. *In vivo*, the SH2 domain of mammalian Grb2 binds phosphorylated sites on activated receptors and cytoplasmic phosphoproteins that conform to this consensus^{19,20,24,25}, whereas its SH3 domains bind Pro-rich motifs in the Ras guanine nucleotide exchange factor Sos^{15,21,22}, thereby mediating Ras activation⁴. To investigate the potential influence of Trp at residue EF1 on SH2 binding specificity, the sequence for the Src SH2 domain was altered to encode Trp instead of Thr at this position. A GST fusion protein containing the mutant Src SH2 domain with the Thr to Trp substitution (T215W) still preferentially selected Glu at the +1 position from a degenerate phosphopeptide library (Fig. 1c). However, in contrast to the wt Src SH2 domain, the T215W Src SH2 preferentially bound Asn as the +2 residue and Pro at the +3 position (Fig. 1c). Hence, changing the EF1 residue in the Src SH2 domain from Thr to Trp alters its principal selectivity from pYEEI to pYENP.

The kinetics of wt or T215W Src SH2 domain binding to specific phosphopeptides containing the sequences pYEEI or pYENP were assessed (Table 1). The wt Src SH2 bound to the pYEEI peptide with a dissociation constant (K_d) of ~80 nM, but about 100-fold less efficiently to the pYENP peptide. In contrast, the T215W Src SH2 domain bound poorly to the pYEEI peptide, but with high affinity to the pYENP peptide (K_d = 110 nM). We also measured binding to a phosphopeptide with the sequence pYVNV, which corresponds to the binding site for the Grb2 SH2 domain on the mammalian Shc and Bcr-Abl proteins^{23,25}. Unlike wt Src SH2, the T215W Src SH2 domain bound strongly to the pYVNV peptide, with an affinity close to that of the Grb2 SH2 domain (Table 1). Substituting the TrpEF1 residue found in the wt Grb2 SH2 domain with Thr (W121T) led to a decrease in its affinity for the pYVNV peptide (Table 1), correlating with a loss of selectivity for the +3 position as determined using a degenerate phosphopeptide library (data not shown).

These data indicate that the variant T215W Src SH2 domain has a modified binding specificity. To pursue this possibility, phosphorylated focal adhesion kinase (FAK) and Shc proteins

^{*} To whom correspondence should be addressed.

were examined for their binding to wt Src, T215W Src and Grb2 SH2 domains. FAK is a 125K tyrosine kinase, concentrated in focal adhesions, that associates *in vivo* and *in vitro* with the Src family tyrosine kinases c-Src and Fyn²⁶⁻²⁸. Binding of Src to FAK is mediated by the Src SH2 domain, which specifically recognizes the major site of autophosphorylation at Tyr 397,

within the motif pYAEI (refs 28, 29; Fig. 2a, lanes 2 and 3). In contrast to the wt Src SH2 domain, the T215W Src SH2 domain bound poorly to autophosphorylated FAK *in vitro* (Fig. 2a, lane 4). Another mutant Src SH2 domain with a substitution of Gly for Thr at the EF1 site, which possesses a similar selectivity to the wt Src SH2 in the degenerate phosphopeptide assay (data

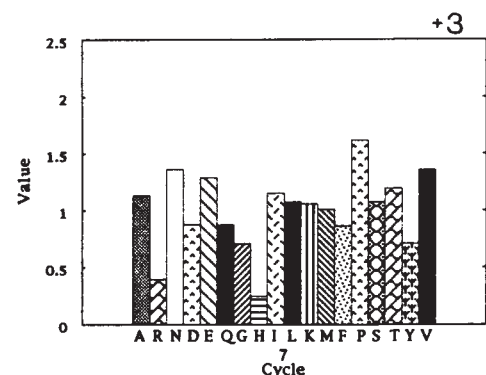
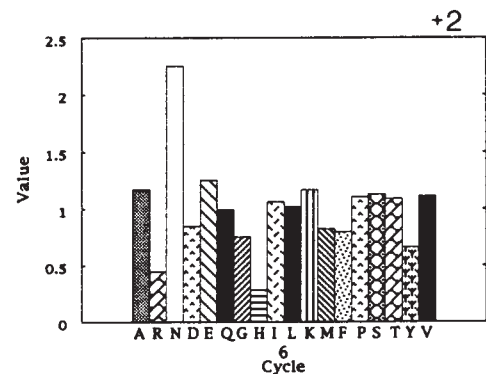
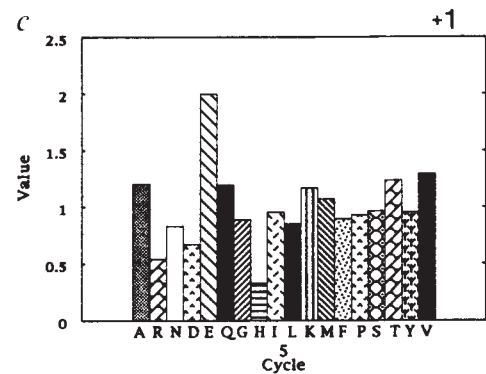
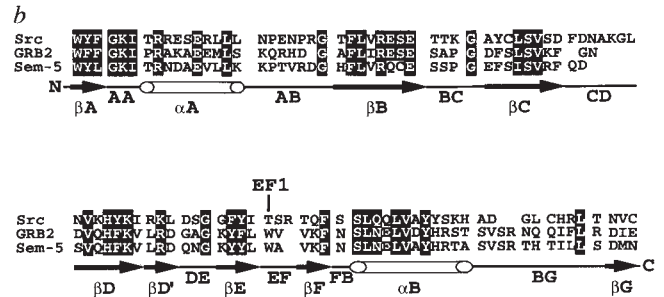
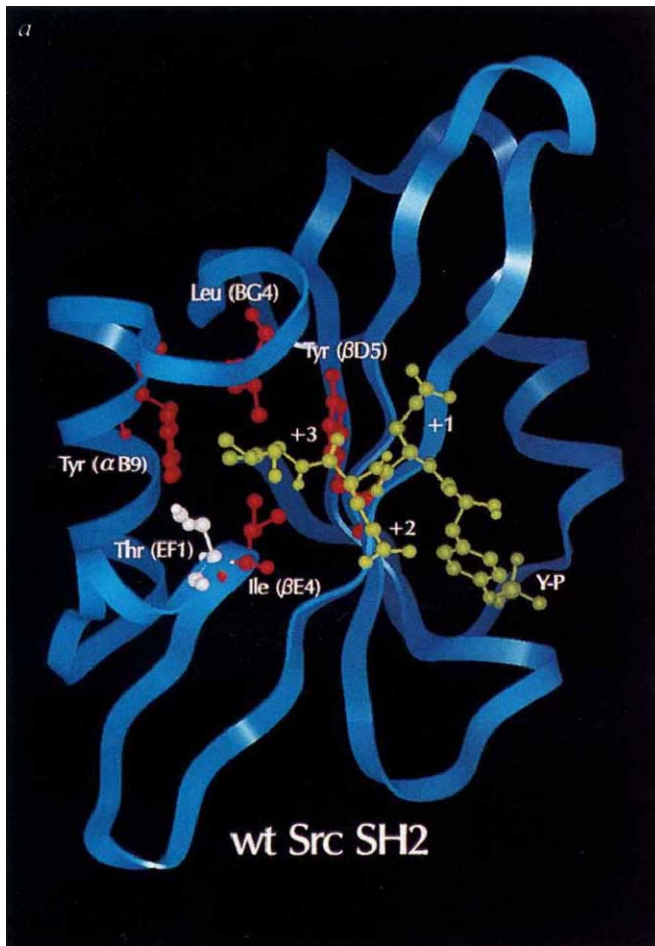


FIG. 1 Phosphopeptide selectivity of the mutant T215W Src SH2 domain. **a**, The X-ray crystallographic structure of the wt Src SH2 domain¹², displayed as a blue ribbon, is shown bound to a pYEEI peptide (in green). ThrEF1 is in white, and other residues forming the binding pocket for the +3Ile are in red. **b**, Sequence alignment of the Src and Sem-5/Grb2 SH2 domains. The locations of the α -helices, β -sheets and loops in the Src SH2 domain are shown. The EF1 residue is indicated. Note that the Src EF loop has three residues, whereas the corresponding element in the Sem-5/Grb2 SH2 domain has only two residues. **c**, A phosphopeptide library, degenerate at the +1, +2 and +3 positions relative to phosphotyrosine, was added to a column of immobilized T215W Src SH2 domain as a GST fusion protein. Bound peptides were eluted and sequenced⁵. Panels show results from the 5th, 6th and 7th cycle of the sequence, representing the +1, +2 and +3 positions C-terminal to the phosphotyrosine. The value represents the amount of each amino acid eluted from the GST-T215W Src SH2 column, relative to the same amino acid from the control GST column. An amino acid which shows no selection will have a value of one or less.

METHODS. The codon for Src Thr 215 (corresponding to the EF1 position of the SH2 domain) was mutated to encode Trp. The mutant T215W Src SH2 domain (containing Src residues 148–252) was sequenced and subcloned into the pGEX4T2 GST bacterial expression vector. Selection and analysis of phosphopeptides by GST-SH2 fusion proteins from a peptide library has been described in detail elsewhere⁵. Peptides in the library have the sequence GDGPYXXXSPLLL, in which pY is phos-

photyrosine and X is a degenerate position containing all amino acids except for Cys or Trp.

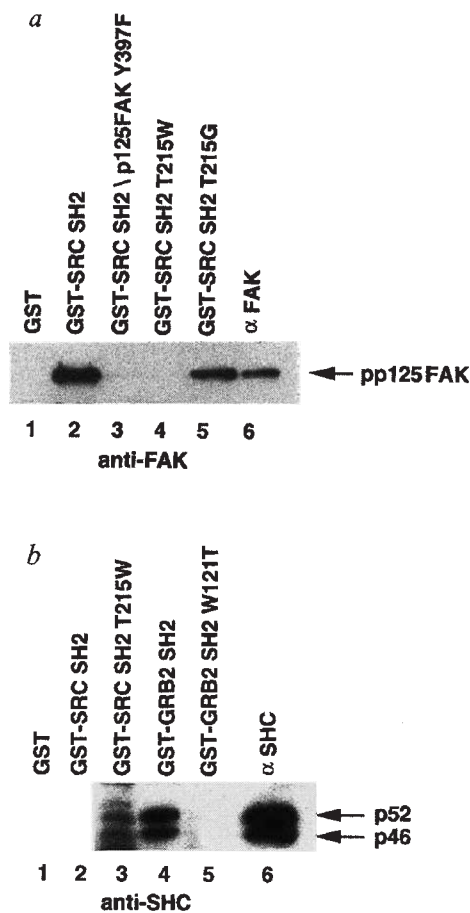


FIG. 2 Binding of mutant Src and Grb2 SH2 domains to FAK and Shc. *a*, Association of SH2 domains with FAK. Chicken embryo fibroblasts (CEF) engineered to over-express wt FAK²⁸ were lysed. Cell lysates were then incubated with immobilized GST fusion proteins containing parental GST (lane 1), wt Src SH2 (lane 2), T215W Src SH2 (lane 4), and T215G Src SH2 (lane 5), or with anti-FAK antibodies (lane 6). CEF overexpressing a FAK mutant Y397F, in which Tyr 397 is substituted by Phe (refs 28, 29), were lysed and incubated with an immobilized GST fusion protein containing wt Src SH2 (lane 3). Complexes were separated by SDS–polyacrylamide gel electrophoresis (PAGE), transferred to a nitrocellulose filter, and immunoblotted with rabbit anti-FAK antibodies, followed by ¹²⁵I-protein A. The mobility of FAK is indicated. *b*, Association of SH2 domains with Shc. *v-src*-transformed Rat-2 cell lysates were incubated with parental GST (lane 1), with GST fusion proteins containing wt Src SH2 (lane 2), T215W Src SH2 (lane 3), wt Grb2 SH2 (lane 4), or W121T Grb2 SH2 (lane 5), or anti-Shc antibodies (lane 6). Complexes were separated by SDS–PAGE and immunoblotted with anti-Shc antibodies followed by ¹²⁵I-protein A. The mobilities of the 46 and 52K Shc isoforms are shown. METHODS. CEF were infected with an RCAS retroviral vector containing the avian FAK cDNA²⁹. CEF and Rat-2 *v-src* cells were grown and lysed as described^{10,29}. Cell lysates (2 mg ml⁻¹) were then incubated with 2 μg of each immobilized GST-SH2 fusion protein in a final volume of 1 ml for 1 h at 4 °C. Complexes were washed and immunoblotted as described^{10,19}. Rabbit anti-Shc and anti-FAK antibodies have been described^{26,29}. The T215G Src SH2 domain has a substitution of Gly for Thr at the EF1 position. The wt Grb2 SH2 domain has residues 59–159 of human Grb2. The W121T Grb2 SH2 domain has a substitution of Thr for Trp at the EF1 position (Grb2 residue 121). GST fusion proteins were expressed in *E. coli*, isolated, and immobilized on glutathione-agarose beads as previously described^{10,19}.

not shown), still bound stably to tyrosine phosphorylated FAK *in vitro*. Hence, the decreased binding of the T215W Src SH2 domain to FAK correlates with its decreased selectivity for the residues found at the +2 and +3 positions of the FAK Tyr 397 autophosphorylation site. The *Shc* gene encodes two prominent proteins of 46 and 52K, that are phosphorylated by a range of tyrosine kinases at Tyr 317

within the sequence pYVNV, and consequently associate with the Grb2 SH2 domain to form a complex implicated in regulation of the Ras pathway^{20,23}. Although the wt Src SH2 domain did not detectably bind phosphorylated Shc proteins, the variant T215W Src SH2 domain bound to the 46 and 52K phosphorylated Shc polypeptides with an efficiency approaching that of the Grb2 SH2 domain (Fig. 2*b*, lane 3). Conversely, the mutant W121T Grb2 SH2 domain, which binds poorly to the pYVNV peptide (Table 1) failed to bind stably to Shc proteins (Fig. 2*b*, lane 5). These results indicate that the T215W Src SH2 domain has been converted to a binding specificity more closely resembling that of Grb2. The *sem-5* gene of *C. elegans* has a genetically defined function in the induction of the hermaphrodite vulva. Mutations in *sem-*

TABLE 1 Kinetic analysis of SH2 domain association with phosphotyrosine-containing peptides

GST-SH2 Fusions	(p)YEEI			(p)YENP			(p)YVNV		
	k_{on} (M ⁻¹ s ⁻¹) 10 ⁵	k_{off} (s ⁻¹) 10 ⁻¹	K_d (M) 10 ⁻⁶	k_{on} (M ⁻¹ s ⁻¹) 10 ⁵	k_{off} (s ⁻¹) 10 ⁻¹	K_d (M) 10 ⁻⁶	k_{on} (M ⁻¹ s ⁻¹) 10 ⁵	k_{off} (s ⁻¹) 10 ⁻¹	K_d (M) 10 ⁻⁶
WT SRC	39.3	3.5	0.08	0.21	1.8	8.57	0.39	0.46	1.2
SRC T215W	2.7	4.2	1.6	13.07	1.5	0.11	16.6	0.34	0.02
WT GRB2	ND	ND	ND	0.25	2.9	11.6	26.0	0.33	0.012
GRB2 W121T	ND	ND	ND	4.4	2.4	0.54	2.1	0.36	0.17

Purified soluble GST fusion proteins containing wt Src, T215W Src, wt Grb2, or W121T Grb2 SH2 domains were analysed for binding to phosphopeptides EPQpYEEIPIYLK, EPQpYENPPIYLK, and DPSPYVNVQNLDK (abbreviated as pYEEI, pYENP and pYVNV, respectively) using surface plasmon resonance (SPR) technology (Biacore, Pharmacia Biosensor). Calculated k_{on} , k_{off} and K_d values are shown for each interaction. Phosphopeptides, as 0.5 mM solutions in 50 mM HEPES, pH 7.5 and 2 M NaCl, were immobilized on the surface of sensor chips following conditions previously described³¹. Purified GST-SH2 fusion proteins at concentrations ranging from 100 to 300 nM in 50 mM HEPES, pH 7.5, 150 mM NaCl and 3.0 mM EDTA (HBS buffer) were injected across the surface and the resonance signal recorded. Independent experiments were done to measure the association (k_{on}) and dissociation (k_{off}) rate constants. The k_{on} values were obtained from the SPR signal binding data and calculated using the BIA-evaluation software (Pharmacia Biosensor). To determine the k_{off} values, a 300 nM solution GST-SH2 domain fusion protein was injected across the phosphopeptide-immobilized sensor chip surface and the SPR signal monitored following injection of excess free phosphopeptide (50 μM). The SPR response at selected time points was fitted to a single exponential decay equation (ENZFITTER, Elsevier-Biosoft) and the k_{off} value calculated. The K_d values were obtained following the equation $K_d = k_{off}/k_{on}$. All kinetic runs were in triplicate; individual numbers given for k_{on} , k_{off} and K_d are mean values.

Microinjected constructs	SH3 SH2 SH3	Rescued worms (non-Vul/total)
wt Sem-5		73% (125/172)
Sem-5/ Src SH2		21% (36/170)
Sem-5/ T215W Src SH2		59% (96/162)
W122T Sem-5		13% (17/129)
	EF1	

FIG. 3 Rescue of *C. elegans* Vulvaless phenotype by the T215W Src SH2 domain. A *sem-5* expression vector was modified to encode either wt Sem-5, Sem-5/Src SH2 in which the Sem-5 SH2 domain is replaced by the Src SH2 domain, Sem-5/T215W Src SH2 in which the Sem-5 SH2 domain is replaced by the mutant T215W Src SH2 domain, or W122T Sem-5 in which the TrpEF1 Sem-5 SH2 residue is substituted by Thr. The structure of each protein is shown, together with the amino acid at residue EF1. Sem-5 sequences are shown in black boxes, and Src sequences in white boxes. DNAs, at $4 \mu\text{g ml}^{-1}$, were co-injected with transformation marker DNA (pRF4 at $100 \mu\text{g ml}^{-1}$) into NH2034 (*clr-1(e1745ts); sem-5(n2030)*), a *C. elegans* strain which displays a Vulvaless phenotype. Rescued animals are non-Vulvaless (non-Vul). Transient rescue assay results are shown. Animals injected with pRF4 alone showed no rescue.

METHODS. Expression vectors were constructed using the vector NH#35. This contains the *sem-5* promoter and genomic coding sequence, with the exception that the genomic region encoding the SH2 domain is replaced by a corresponding *Kpn1-Nhe1* fragment from the *sem-5* cDNA. The sequence in NH#35 encoding Sem-5 SH2 (Sem-5 residues 62–153) was replaced with a sequence for the wt Src SH2 (Src residues 151–244) or the T215W Src SH2, using recombinant PCR techniques. The two N-terminal SH2 residues, which are identical in Src and Sem-5, are still encoded in each case by *sem-5* sequences. To create a 5' *Kpn1* site in the Src SH2 domain coding sequence to aid subcloning, a mutation was introduced that alters Phe β A3 to Leu, as found in the Sem-5 SH2 domain. No effect of this substitution alone was observed on *in vitro* binding of the Src SH2 domain to phosphotyrosine-containing proteins. All constructs were sequenced. The Sem-5 W122T mutation was engineered into the *sem-5* cDNA by PCR, and subcloned into the NH#35 vector. Transformation of *C. elegans* and scoring of the Vulvaless/non-Vulvaless phenotypes have been previously described^{14,19}.

5, including those which affect SH2 residues implicated in binding to phosphotyrosine, cause a decreased activation of the vulval inductive pathway^{14,19}. The resulting Vulvaless phenotype is apparently due to an inability of the Let-23 receptor tyrosine kinase to activate the Ras pathway in the absence of functional Sem-5 protein. To test the biological importance of the EF1 residue in the Src and Sem-5/drk/Grb2 SH2 domains, the coding sequence for the Sem-5 SH2 domain was excised and replaced with the sequence for the Src SH2. The resulting construct, which contains the *sem-5* promoter, should direct synthesis of a chimaeric protein (Sem-5/Src SH2) which is identical to wt Sem-5 in overall structure, but contains the Src SH2 domain instead of the Sem-5 SH2 domain. This construct, and a construct encoding wt Sem-5 protein, were microinjected into a *C. elegans* strain which carries a strong *sem-5* allele (*n2030*), and displays a completely penetrant Vulvaless phenotype¹⁴. Transiently transformed progeny of these microinjected animals were scored for suppression of the Vulvaless phenotype (Fig. 3). Although wt *sem-5* restored a normal vulva in 73% of microinjected worms, the Sem-5/Src SH2 construct rescued only 21%. This defect in vulval induction is presumably due to a decrease in binding affinity for the autophosphorylated Let-23 receptor tyrosine kinase or an intermediate phosphoprotein.

A sequence encoding the variant T215W Src SH2 domain was then introduced into the *sem-5* construct, which therefore codes

for a protein in which the T215W Src SH2 domain is surrounded by the Sem-5 SH3 domains (Sem-5/T215W Src SH2). This construct restored vulval induction to a level approaching that of wt *sem-5* (Fig. 3). Substitution of ThrEF1 of the wt Src SH2 domain with Trp therefore enhanced the ability of the Src SH2 domain to replace the Sem-5 SH2 domain in a biological setting by about threefold. In a complementary experiment, the TrpEF1 residue of wt Sem-5 (residue 122, homologous to Grb2 residue 121) was changed to Thr, a substitution which decreases the ability of the homologous Grb2 SH2 domain to bind phosphorylated Shc proteins and the pYVNV peptide (Fig. 2b, Table 1). This W122T Sem-5 mutant had markedly decreased biological activity. Hence, the amino acid at SH2 position EF1 is important for the proposed adaptor function of Sem-5 in coupling the Let-23 tyrosine kinase to the Ras pathway. The results of transient assays were corroborated in each case by the generation of stable lines (data not shown).

Structural analysis has indicated that residues in the EF and BG loops are important in determining SH2 specificity^{12,13,30}. Molecular modelling of the Src SH2 domain containing Trp instead of Thr at residue EF1 indicates that the Trp indole ring could protrude into the +3 pocket, and impose selection for amino acids with smaller sidechains, such as Pro or Val. However, selectivity for Asn at the +2 position suggests that TrpEF1 may induce a more complex change in the orientation of the EF loop. Determining the precise role of the TrpEF1 residue in ligand binding will require direct structural analysis. Experimentally, however, we have found that the mutant T215W Src SH2 domain exhibits a binding specificity towards phosphorylated peptides and proteins resembling that of Sem-5/drk/Grb2. This altered binding specificity is accompanied by an increased ability of the T215W Src SH2 domain to substitute for the Sem-5 SH2 domain in determining vulval cell fate in *C. elegans*. These results are consistent with the view that SH2 binding specificity determines the ability of cytoplasmic signalling proteins to couple specific tyrosine kinases to downstream biochemical pathways. □

Received 26 January; accepted 14 April 1994.

- Anderson, D. et al. *Science* **250**, 979–982 (1990).
- Pawson, T. & Gish, G. *Cell* **71**, 359–362 (1992).
- Cantley, L. C. et al. *Cell* **64**, 281–302 (1991).
- Pawson, T. & Schlessinger, J. *J. curr. Biol.* **3**, 434–442 (1993).
- Songyang, Z. et al. *Cell* **72**, 767–778 (1993).
- Matsuda, M., Mayer, B. J., Fukui, Y. & Hanafusa, H. *Science* **248**, 1537–1539 (1990).
- Fanti, W. J. et al. *Cell* **69**, 413–423 (1992).
- Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. *Science* **252**, 668–674 (1991).
- Mayer, B. J., Jackson, P. K. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 627–631 (1991).
- Marengere, L. E. M. & Pawson, T. *J. biol. Chem.* **267**, 22779–22786 (1992).
- Waksman, G. et al. *Nature* **358**, 646–653 (1992).
- Waksman, G., Shoelson, S., Pant, N., Cowburn, D. & Kuriyan, J. *Cell* **72**, 779–790 (1993).
- Eck, M. I., Shoelson, S. E. & Harrison, S. C. *Nature* **362**, 87–91 (1993).
- Clark, S. G., Stern, M. J. & Horvitz, H. R. *Nature* **356**, 340–344 (1992).
- Olivier, J. P. et al. *Cell* **73**, 179–191 (1993).
- Simon, L. A., Dodson, G. S. & Rubin, G. M. *Cell* **73**, 169–177 (1993).
- Lowenstein, E. J. et al. *Cell* **70**, 431–442 (1992).
- Songyang, Z. et al. *Molec. cell. Biol.* **14**, 2777–2730 (1994).
- Stern, M. J. et al. *Molec. Biol. Cell* **4**, 1175–1188 (1993).
- Rozakis-Adcock, M. et al. *Nature* **360**, 689–692 (1992).
- Rozakis-Adcock, M., Fernley, R., Wade, J. P., Pawson, T. & Bowtell, D. *Nature* **363**, 83–85 (1993).
- Li, N. et al. *Nature* **363**, 85–88 (1993).
- Pellicci, G. et al. *Cell* **70**, 93–104 (1992).
- Pendergast, A. M. et al. *Cell* **75**, 175–185 (1993).
- Puil, L. et al. *EMBO J.* **13**, 764–773 (1994).
- Schaller, M. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5192–5196 (1992).
- Hanks, S. K., Calalb, M. B., Harper, M. C. & Patel, S. K. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8487–8491 (1992).
- Cobb, B. S., Schaller, M. D., Leu, T.-H. & Parsons, J. T. *Molec. cell Biol.* **14**, 147–155 (1994).
- Schaller, M. D. et al. *Molec. cell Biol.* **14**, 1680–1688 (1994).
- Pascal, S. et al. *Cell* **77**, 461–472 (1994).
- Felder, S. et al. *Molec. cell Biol.* **13**, 1449–1455 (1993).

ACKNOWLEDGEMENTS. We thank J. Kuriyan for the coordinates for the Src SH2 domain, J. Schlessinger for the GRB2 cDNA, J. Bell for synthesis of phosphopeptides and M. Berne for peptide sequencing. Z.S. thanks EDS for technical assistance. This work was supported by grants from the National Cancer Institute of Canada (NCIC), and the Medical Research Council of Canada and by a Howard Hughes International Research Scholar Award to T.P., and by grants from the Lucille P. Markey Charitable Trust to L.C.C. and M.J.S. J.T.P. and M.D.S. are supported by the DHHS. T.P. is a Terry Fox Cancer Research Scientist of the NCIC.

Allergy by mutation

The search for genetic factors that may predispose people to various allergies takes a promising turn with the publication of some incriminating evidence against the IgE receptor.

FOR many people, common complaints such as hay fever, eczema and food allergies are little more than an irritating nuisance. But at the extreme end of the allergy spectrum, asthma results in thousands of deaths each year. The causes of these conditions, which are given the collective term 'atopy', are thought to involve the familiar conspiracy between environmental and genetic factors. The environmental culprits are fairly well known, including pollens and house dust mites. When these allergens bind to IgE, which is itself held captive on the surface of mast cells by its receptor, it triggers the release of inflammatory substances, such as histamine, which lead to the familiar spring symptoms — sneezing, coughing, wheezing and so on. In atopic individuals, the rise in serum levels of IgE in response to allergen exposure is both prolonged and considerably enhanced.

But what about the genes that might make some individuals particularly susceptible to atopic allergies? A report by Julian Hopkin, William Cookson and colleagues in Oxford, in the June issue of *Nature Genetics*, turns the spotlight onto what in hindsight is a rather attractive candidate gene — that for the high-affinity IgE receptor (FcεRI)¹.

For the past five years, Hopkin and colleagues have been accumulating evidence that a gene on chromosome 11 is associated with up to 60% of atopy cases^{2,3}. Prompted by earlier suggestions that the risk of atopy is higher in the children of atopic mothers than affected fathers, they also found that the inheritance of atopy at this locus occurred exclusively through the maternal line³. Attempts to replicate these findings have met with mixed results⁴, although two other recent studies have supported the Oxford group's conclusions.

Last year, Hopkin and colleagues refined the critical atopy region to the point where they could start testing potential candidate genes⁵. After that encoding

CD20 proved negative, they turned to a homologous gene encoding the β-subunit of FcεRI. They first sequenced the gene from six atopic individuals, and in one of them found a mutation in one of the transmembrane domains of the receptor: a substitution of a leucine residue for isoleucine at position 181.

Next, they examined the prevalence of the mutation in two populations. A screen of 163 people selected at random turned up the Ile181Leu variant in 25 (15%), 13 of whom were found to be atopic (as measured by total IgE levels and reaction to grass pollen antigen). Given that earlier studies suggested the transmission was parentally influenced, it may not be surprising to see the disorder in only half of the subjects bearing the putative errant allele. They also looked at 60 nuclear families with atopy, and encountered the same mutation in ten of them. In each case, the mutation was maternally inherited, and all 12 children that had inherited Leu181 had developed marked atopy.

It is possible that the Ile181Leu variant is simply in linkage disequilibrium with another, unidentified mutation at the FcεRI-β locus, but there are reasons to think otherwise. The β-chain of FcεRI is the least well understood subunit of the heterotrimeric IgE receptor; nonetheless, although it does not seem to be required for expression or crosslinking of the molecule, new findings suggest that it has an important and previously unsuspected role in autophosphorylation and signal transduction⁶. Furthermore, a similar mutation in a homologous receptor (the ζ-chain of the IgG Fc receptor, FcγRIII) has been shown to affect receptor assembly⁶. It should be possible, therefore, to examine directly the effect of the Ile181Leu variant on FcεRI-β function.

Good evidence for the existence of genetic heterogeneity in atopy was presented last month by researchers at Johns Hopkins, who performed sib-pair analysis on families in the Amish population and, using a candidate gene approach, mapped a locus on the long arm of chromosome 5 that seems to control total IgE serum levels⁷. Interestingly, this region contains a cluster of several cytokine genes, including that encoding interleukin-4, which plays an important role in T-cell differentiation and IgE production by B lymphocytes. Thus, both candidates identified so far are perfectly plausible, and efforts are continuing to identify other loci involved in atopy. Sorting out the

relative contributions of these genes will doubtless be a story for another day.

Since 1989, when the gene for cystic fibrosis (CF) was first identified, more than 400 different mutations have been documented, although the first of them — the deletion of a single phenylalanine residue at position 508 (ΔF508) — is far and away the most common, accounting for 70% of all CF chromosomes. But that figure conceals a stark and surprising variation within Europe: in the north of Europe and Scandinavia, the prevalence of ΔF508 is as high as 90%; but around the Mediterranean, it makes up less than half of the total mutations. So why has ΔF508 assumed such a high overall frequency and graded distribution?

Another study in the June issue of *Nature Genetics*, one coordinated by Xavier Estivill in Barcelona, and involving workers from 13 different European countries, has tackled just this question⁸. The group looked at more than 1,700 CF chromosomes bearing the ΔF508 mutation, and catalogued the haplotype of each chromosome, as composed of three highly polymorphic markers within the CF gene. Of a total of 54 different haplotypes with these three markers, just four accounted for nearly 90% of the chromosomes. By assessing the variability, or 'slippage', at each marker with respect to each other, the team were able to calculate the original haplotype on which ΔF508 was thought to have arisen, and measure the expected time required to account for the variability seen today.

That result turns out to be in excess of 2,500 generations, which (allowing 20 years per generation) yields an origin for ΔF508 of more than 50,000 years ago — considerably older than previous estimates. The authors suggest that the common CF mutation arose in an ancestral European population, and later spread through two distinct populations. The remarkable persistence of ΔF508 supports the notion that CF carriers have a heterozygous advantage, perhaps against cholera or diarrhoea.

Kevin Davies

Kevin Davies is the editor of Nature Genetics.

1. Shirakawa, T. et al. *Nature Genet.* **7**, 125–130 (1994).
2. Cookson, W.O.C.M. et al. *Lancet* **i**, 1292–1295 (1989).
3. Cookson, W.O.C.M. et al. *Lancet* **340**, 381–384 (1992).
4. Marsh, D.G. & Meyers, D. *Nature Genet.* **2**, 252–254 (1992).
5. Sandford, A.J. et al. *Lancet* **341**, 332–334 (1993).
6. Ravetch, J.V. *Nature Genet.* **7**, 117–118 (1994).
7. Marsh, D.G. et al. *Science* **264**, 1152–1156 (1994).
8. Morral, N. et al. *Nature Genet.* **7**, 169–175 (1994).

Also in June's *Nature Genetics*: different subunits of the glycine receptor mutated in *spastic* and *spasmodic* mouse strains; complete map of the immunoglobulin V_H region on chromosome 14; African origins of oculocutaneous albinism; linkage mapping of Sorsby's fundus dystrophy; and the identification of novel bacterial transcripts by sequence analysis.